Towards the development of a genetic test for distinguishing the Hereford Original Population (Hereford (OP)) from the Hereford North American Derived (Hereford (NAD)) cattle in the UK, Australia and North America

Hegarty, Matthew; Broadbent, Zoe Elizabeth; McMahon, Robert

DOI:
10.20391/dhkb-vh60

Publication date:
2020

Citation for published version (APA):
Hegarty, M., Broadbent, Z. E., & McMahon, R. (2020). *Towards the development of a genetic test for distinguishing the Hereford Original Population (Hereford (OP)) from the Hereford North American Derived (Hereford (NAD)) cattle in the UK, Australia and North America*. Prifysgol Aberystwyth | Aberystwyth University. https://doi.org/10.20391/dhkb-vh60

PRIFYSGOL ABERYSTWYTH UNIVERSITY

IBERS
Athrofa y Gwyddorau Biolegol, Amgylcheddol a Gwledig
Institute of Biological, Environmental and Rural Sciences

Report prepared for Hereford Herd Book International Ltd – October 2020

Towards the development of a genetic test for distinguishing the Hereford Original Population (Hereford (OP)) from the Hereford North American Derived (Hereford (NAD)) cattle in the UK, Australia and North America

Matt Hegarty, Zoe Broadbent and Rob McMahon

Final report copy for distribution

Dr. Matt Hegarty
11th October 2020

1.1. Executive Summary

The purebred Hereford Original Population (Hereford (OP)) cattle breed – sometimes known as the Traditional Hereford - has existed in the UK since the mid-1800s and is the forerunner of the Hereford North American-Derived (Hereford (NAD)) which has diverged genetically as a result of hybridisation and is largely derived from breeding stock developed in North America following colonisation and continues to the present day. The Hereford (NAD) is a longer-legged, larger skeletal framed animal resulting from breeding in North America and subsequently re-imported into the UK, Australia and the rest of the world. Known for its hardiness, the shorter-legged Hereford (OP) has Category 5 rare breed status within the UK (~1500 breeding cows). The Rare Breed Survival Trust (RBST) and Hereford Herd Book International Ltd (HHBI) commissioned this report in order to protect the breed both commercially through meat provenance testing and genetically via prevention of admixture with external bloodlines. The Hereford (OP) breed was also widely used in Canada, South America and Australia, where it was favoured due to its resilience, and Hereford (OP) cattle still present in those regions also require provenance testing.

Previous studies have demonstrated that the two groups are genetically distinct on a scale that indicates introgression of genetic material from one or more additional breeds into the Hereford (NAD). To prevent the risk of outside genetics entering the Hereford (OP) gene pool, a large amount of genotype data has been generated through high-density genetic marker testing of Hereford (OP) animals from the UK, Australia and the USA, representing bloodlines dating back, in some cases, to the 1820s. This project seeks to utilise this data, together with other breed information, to develop a test for distinguishing Hereford (OP) from Hereford (NAD), and to establish the genetic continuity of the Hereford (OP) animals and their distinctiveness from the Hereford (NAD). We demonstrate that this test can correctly identify UK and Australian animals of known/suspected Hereford (NAD) pedigree and those of known Hereford (OP) pedigree. The test can also be used to assist breeders to breed back towards the purebred Hereford (OP), correctly inform new entrants regarding the genetics of the breed and educate the public in general about the purebred Hereford.

1.2. Use of genotype testing for breed verification in cattle

In a previous project at Aberystwyth University's (AU) Institute of Biological, Environmental and Rural Sciences (IBERS), Dr. Rob McMahon and Dr. Matt Hegarty undertook the development of a carcass verification test with Waitrose and Dovecote Park Farms. As Waitrose's commercial beef supplies range from purebred to quarterbred Angus-sired individuals, the questions were 1) could we confirm parentage by the Angus sire? 2) if that parent was NOT the supposed sire, could we confirm parentage by a different Angus individual? 3) could we confirm the genetic components of the dam? and finally 4) could we potentially assign an animal to region-of-origin or even farm-of-origin?

In order to address these questions, we collected samples from at least 20 representative pedigree sires from each of the major UK beef breeds, in the 2014 cohort. These were examined using a commercially-available genetic test – the Illumina beadchip. These beadchips (and other similar assay systems) rely on genetic markers known as single nucleotide polymorphisms (SNP). Briefly, the DNA of any

individual is made up of several million chemical “letters” (A, T, G and C) and natural variation can cause changes at each of these letters. However, each individual also has two copies of each chromosome (one from each parent), and thus carries two copies of each letter. So, take an example of letter 88,131,171 on human chromosome 4. This can be either A or C. Individuals with two copies of A (AA) have no issues. One (AC) or two (CC) copies of C cause increased risk of gout (the primary author being AC…). Because the SNP only has two possible states, Illumina chips use a two-colour dye test to determine which genotype a given individual has at any given SNP, across thousands or tens of thousands of SNP simultaneously. In our carcass verification work, we employed the BovineLD assay, which genotypes the animals across ~7000 SNP. We also downloaded data from a previously published assay of cattle breeds (Decker *et al.,* 2014) using a larger-scale assay (the Illumina Bovine50K assay, of which the LD test is a subsampling). These datasets were used to demonstrate that the LD test would give the same possibilities of breed identification to the 50K assay, then combined into a single dataset for generation of a breed exclusion test. This project concluded with the ability to identify the breed components of quarterbred animals with 98% accuracy in the majority of cases, using blind-tested samples of known provenance supplied by Dovecote Park Farms (McMahon *et al.*, 2015).

Breed assignment is possible in cattle because the populations have been maintained as effectively closed breeding pools subject to different selective pressures over time. Assume we are using a single genetic marker with two possible genotypes (i.e. one SNP) equally represented in the ancestral population – if there is no selection within the populations, then the overall frequency of both genotypes will remain 50/50, though generation to generation random sampling will result in a general drift between populations. In one population, however, this marker may lie in/near a gene under selection. Over time, one genotype will be co-selected with this gene variant, and thus the frequency becomes skewed (i.e. 80:20). In highly selected systems like cattle, these genotype frequencies can be heavily skewed from one breed to another simply as a result of high parental variance in productivity resulting in overrepresentation of certain particular ancestral genotypes in each derived population. Across enough markers, then, we can detect “signatures” of genotype frequencies that are characteristic of each derived population and no other. In practice, the test works to exclude an individual from each breed, until it is left with a breed (or breeds) from which it cannot be statistically excluded.

1.3. Genetic division within the Herefords

Previous studies by Prof. Jerry Taylor (University of Missouri) and Dr. Rob Ogden (TRACE Wildlife Forensics) have looked at the use of SNP markers to distinguish cattle breeds and understand breed structure and relationships. In particular, Ogden used a set of 96 SNP markers derived from 50K testing to develop an FAO-approved panel for breed verification/exclusion in the meat industry. This test compared unknown samples to reference data collected from animals of known provenance for each breed. In the case of Hereford cattle, discrepancies were identified and studied further in two DEFRA reports (FA0112 and FA0125). To quote from the conclusions of FA0112 (Ogden, 2012):

“The Hereford cattle had a similar issue; however in this case the authenticated positive control samples were not assigned back to the reference population during

validation. The reason for this was considered to be that the reference data was taken from US Commercial Hereford cattle, while the control samples were from UK Traditional Hereford cattle. 'Traditional' and 'Commercial' Hereford in the UK are recognised differently by different cattle stakeholder groups; regardless of the definitions, it was apparent that the two groups were genetically distinct and that genetic diversity within the 'Traditional' Hereford required representation in the reference data; samples were therefore sought to do this.

…

The results for Traditional Hereford samples showed that this group formed a distinct genetic population to the Commercial Hereford originally used as Hereford reference data. The Traditional Hereford samples were therefore treated as a separate 'breed' for the purposes of assignment. All but one of the Traditional Hereford samples was assigned to Traditional Hereford rather than the Commercial Hereford or any other cattle breed. A single sample was excluded from all breeds. In addition, the Traditional Hereford samples from the original project were also assigned to the newly created Traditional Hereford reference population. These results support the establishment of a separate reference population for Traditional Hereford and highlight that the use of US derived population data alone was inadequate and not representative of the meat sold under the Hereford breed label in the UK."

FAO125 largely dealt with the transfer to a new genotyping technology but also expanded the reference set for Traditional Herefords (as Ogden referred to Hereford (OP) animals). It was observed that mis-assignment rates for Herefords were higher than for other breeds, with scope for mis-assignment of modern Hereford to Traditional. Ogden also notes (emphasis ours) that:

"Traditional Hereford and Welsh Black breeds, for example, can be segregated from other breeds using relatively few SNPs which appear either fixed in the reference dataset (i.e. have a single genotype across the entire breed) or exhibit large differences in their allele frequencies when compared to other breeds. Conversely, other breeds such as Commercial Hereford and Red Poll <u>are either recently derived from other breeds or have routinely been crossed in the past with other breeds.</u> This means that they are genetically less distinct from other breeds in the reference dataset and therefore are more difficult to characterise using a small number of SNP markers. Interestingly, when Commercial Hereford individuals are compared to the reference dataset, the majority of those mis-assigned are placed within the Traditional Hereford group. This is likely to reflect the close breeding history of the two breeds and the ancestral origin of Commercial Herefords."

Ogden's use of "Commercial Hereford" is somewhat unfortunate as it implies the Traditional breed is not used commercially, which it is. Also, the term "commercial Hereford" has different connotations in the USA, where it refers to unregistered crossbred animals. Conversely, all of the animals used for our testing here are registered and thus considered "purebred" by the American Hereford Association, Hereford Cattle Society and Herefords Australia. **To avoid confusion, we have adopted the term Hereford (NAD) for these samples, whilst the Traditional Hereford are termed as Hereford (OP).**

Ogden's purpose in development of these tests was to combat fraud in the meat industry – to that end, being able to assign samples to "Hereford" or "Not Hereford" was considered sufficient and ability to distinguish Hereford (OP) and Hereford (NAD) a nice, though unutilised addition. From a breed standard perspective, however, there are several issues arising from this. As Ogden notes, the Hereford (NAD) is genetically clearly distinct from the Hereford (OP) which must be considered the root source, and this is due to the Hereford (NAD) being *"either recently derived from other breeds or have routinely been crossed in the past with other breeds".*

This is emphasised by the phenotypic differences observable between the Hereford (NAD) and Hereford (OP), which are highly unlikely to result from selective breeding from the Hereford (OP) source population, particularly when considering a large jump in body mass in the Hereford (NAD) around the 1990s. The Hereford (NAD) animals are, on average, longer-legged, larger framed and ~200kg heavier. The Hereford (OP) also produces a completely different type of milk to the NAD, being higher in butterfat percentage and total solids *per se*.

The history of breeding in North America strongly supports introgression of material from other breeds at some stage. More recently, the whiteface trait historically associated as the "mark" of the Hereford breed has come under pressure in American killing lines that process branded Hereford beef, with animals being accepted as long as they have white markings on >50% of the face – as a dominant trait believed to be fixed in the Hereford breed, the Hereford whiteface should pass on in full. Evidence has been coming to light (although challenged) that some registered Herefords are no longer consistently passing the whiteface trait to their offspring, which has economic consequences as buyers will not automatically accept such progeny as Hereford crosses. The breakdown of this heritability furthers the case for introgression from other breeds with different spotting mutations, such as Simmental.

If this is the case, the Hereford (NAD) are essentially **a new, composite breed** and common registration of the Hereford (NAD) with the Hereford (OP) breed runs the risk of further introgression of genetic material from other breeds into the UK Hereford (OP) population. The Hereford (OP) is also used further afield, with populations exported to the Americas and Australia either historically or through use of AI/ET. Ensuring that material from the Hereford (NAD) or other crossbred animals does not mix with these Hereford (OP) populations (or detecting where it already has) will require genetic testing. Ogden's test was focused on affordability and thus used a very small SNP set, leading to the possibility of mis-assignment between Hereford (NAD) and Hereford (OP). With more extensive use of high-density SNP arrays, it should now prove possible to distinguish the two breeds with greater accuracy and enable creation of separate registries.

1.4. Aims and Objectives

Using the prior collection of data from our carcass verification work and from Decker *et al.* (2014), together with the new genotypes provided by Hereford Herd Book International (HHBI), IBERS undertook to 1) explore the possibility of a test for identification of Hereford (OP) vs Hereford (NAD), 2) identify which genetic markers strongly differentiate the breeds for commercialisation of this test and 3) determine the degree of breed admixture.

2. Clustering of genotypes

Data were provided from HHBI at various stages during the project, based on commercial genotyping of animals by GeneSeek/NeoGen using the Illumina Infinium assay technology, either in 50K or GeneSeek's custom 150K format. In addition, a further set of data was generated using the 7K BovineLD assay at IBERS from samples collected from the UK Baytal Hereford (OP) herd and collected as part of a M.Sc. project by the student Zoe Broadbent. GeneSeek data was received in or converted to PLINK format (two files: .MAP containing a list of all SNP markers used and their genomic locations, and .PED containing a horizontal list of genotypes for each animal, plus family and individual IDs). In some cases, the datasets for individuals were not in Final Report format or PLINK format and were not able to be included, as to do so would have required extensive (and unavailable) computational support to develop scripts to reformat the data. GeneSeek/NeoGen have been informed of these issues and subsequent commercial testing will always output data in PLINK format for supply to HHBI.

Once in PLINK format (Purcell *et al.*, 2007), the individual files were merged with the dataset from the previous carcass verification project, resulting in a final merged set of 4593 SNP without extensive missing data. To reduce processing load, SNP genotype calls were also converted to binary 1/2 format (for example, an A/G marker would be listed as 1/1 for an AA individual, 1/2 for AG and 2/2 for GG). Following data conversion, a relationship matrix based on shared genotypic Identity-By-State (number of markers showing the same genotype between pairs of animals) was created and Multidimensional Scaling (MDS) analysis used to identify the major sources of variation within the data. These data are output as an Excel-readable file and were used to produce scatterplots showing where individual animals lie in relationship to each other, based on their MDS coordinates. Detailed methodologies used and protocols for these tests are available on request

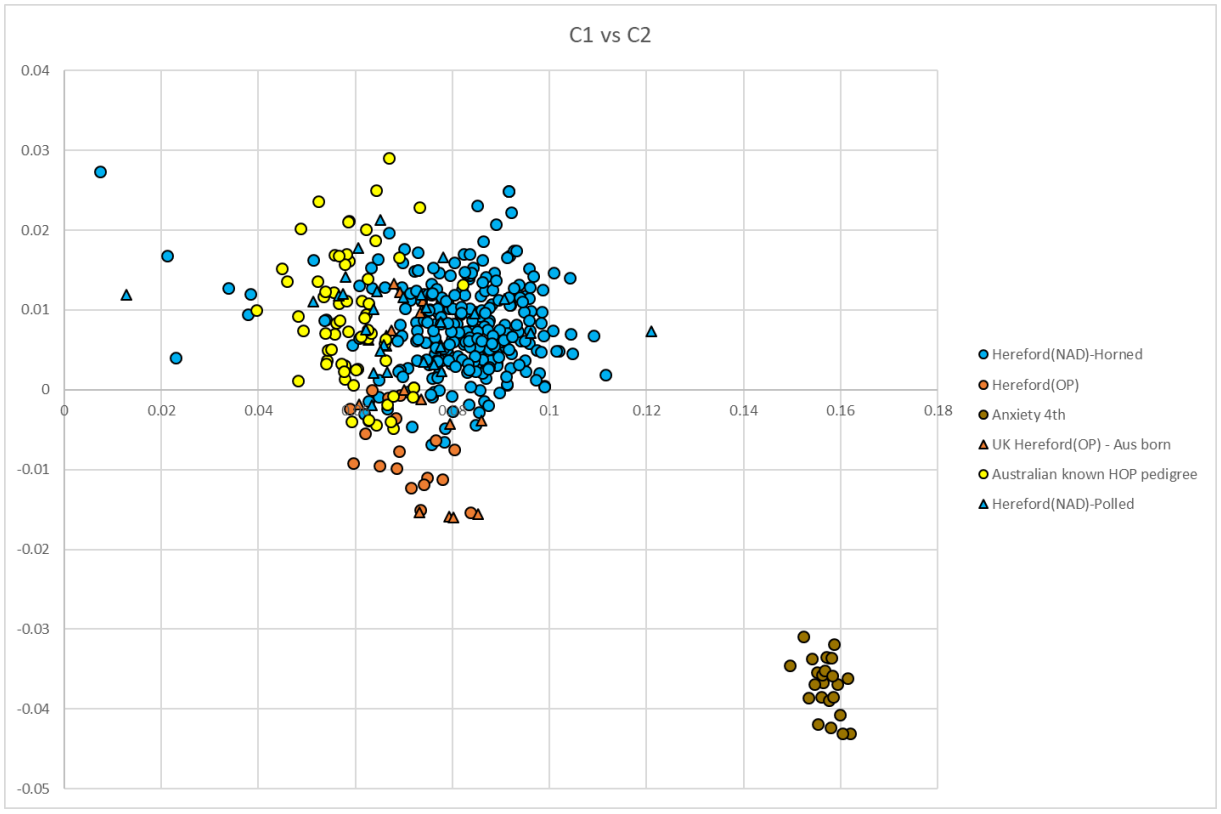
In addition to the samples supplied by HHBI, the data included 1057 animals of other pure breeds or hybrid animals tested as part of the carcass verification work, plus data from Decker *et al*. (2014). The initial reference set of Hereford (OP) animals provided for this were as follows – these animals are the same as supplied to Ogden for FAO112):

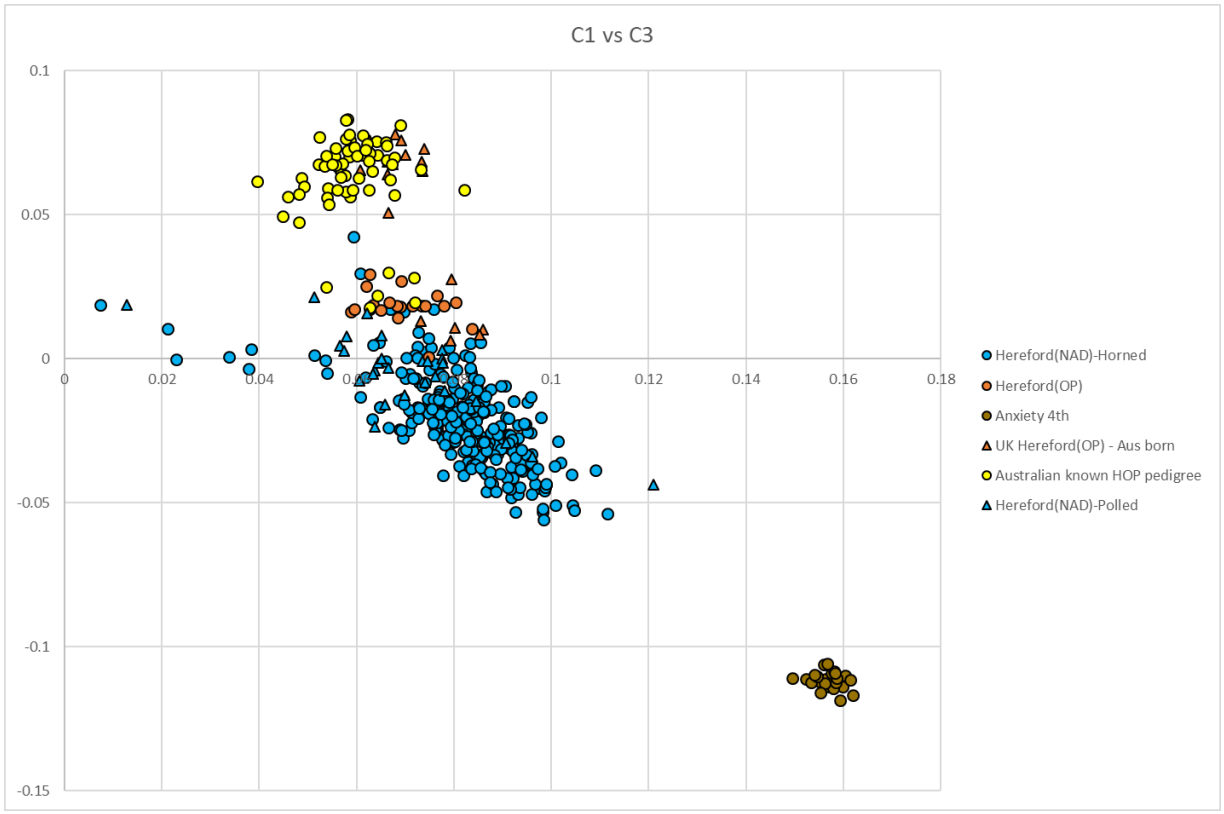
Baytal Countess Harriet
Baytal Amorous Celia
Darmon Fennel
Dornley Dowager 6th
Brougham Aga
Rowley Rebel
Lowerhill Winny
Freeby Laird
Moor Blatant
Avon Claudio
Withington Hero
Bodenham Christopher
Penmaes Kruger
Sutton Marcus
Huddleston Earl
Ffostill Brigadier
Broadway Wonder
North Woodloes Excelsior
Cefnhiniog Wisdom

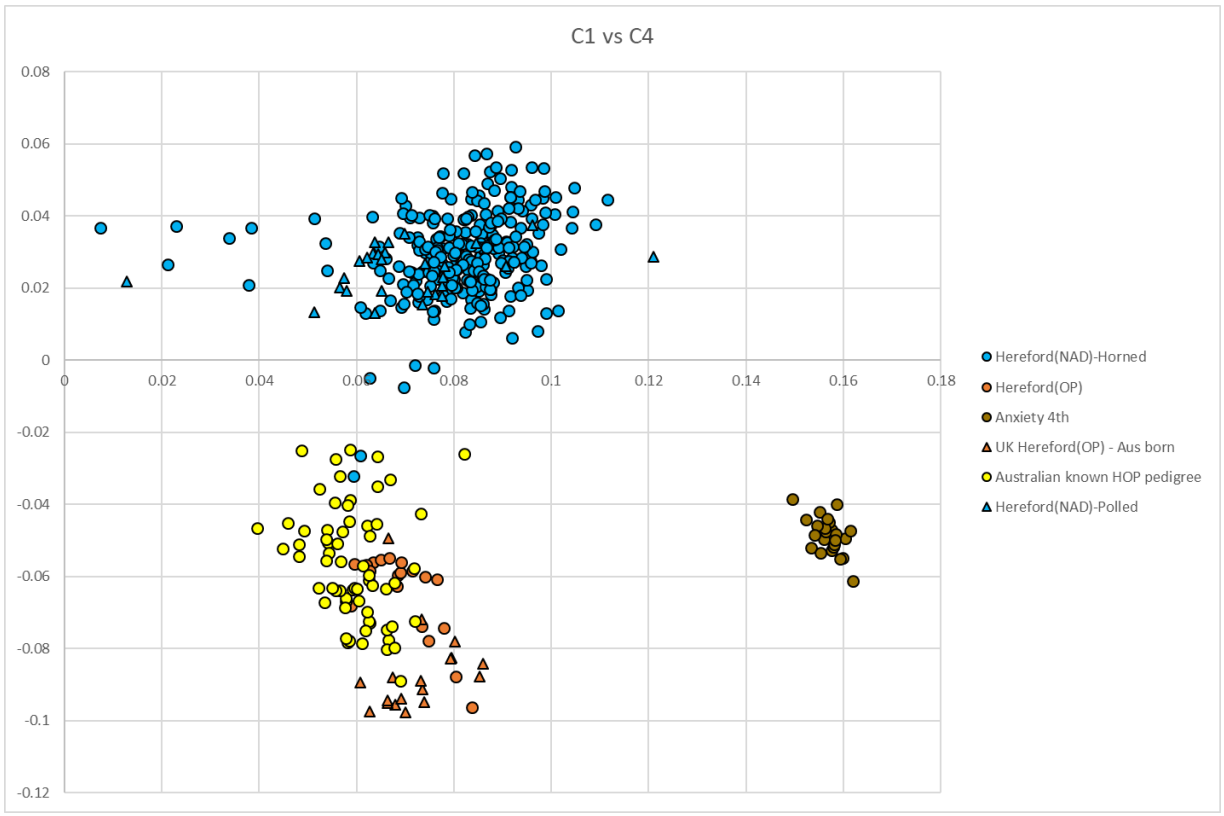
2.1. Results of MDS clustering for Hereford (OP) and Hereford (NAD) reference sets

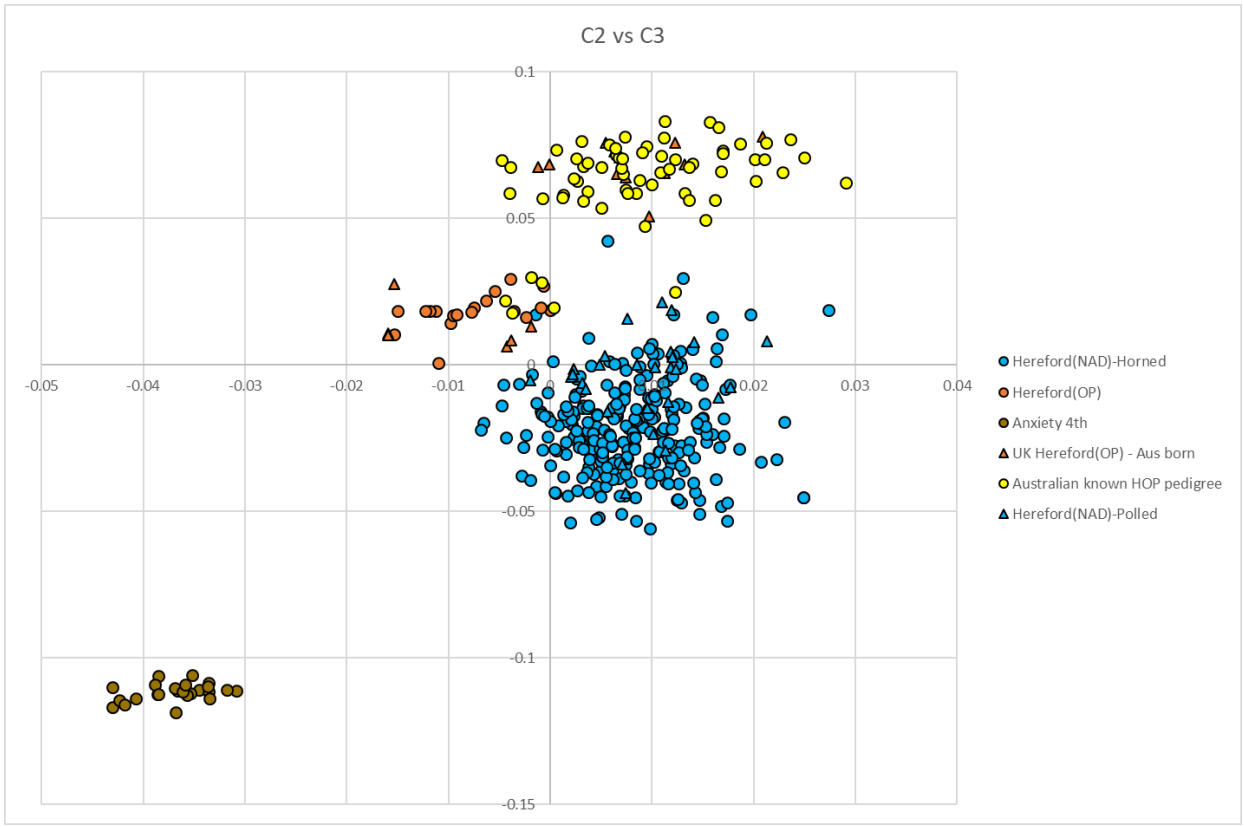
Initial analysis using PLINK confirmed the findings of Ogden and Taylor, showing a clear division between Hereford (OP) and Hereford (NAD) with MDS clustering. The plots below show the clustering of animals based on their assignment to the differing pairs of MDS coordinates, in order to determine that any pattern observed was not solely an artefact of looking at only the first two clusters identified.

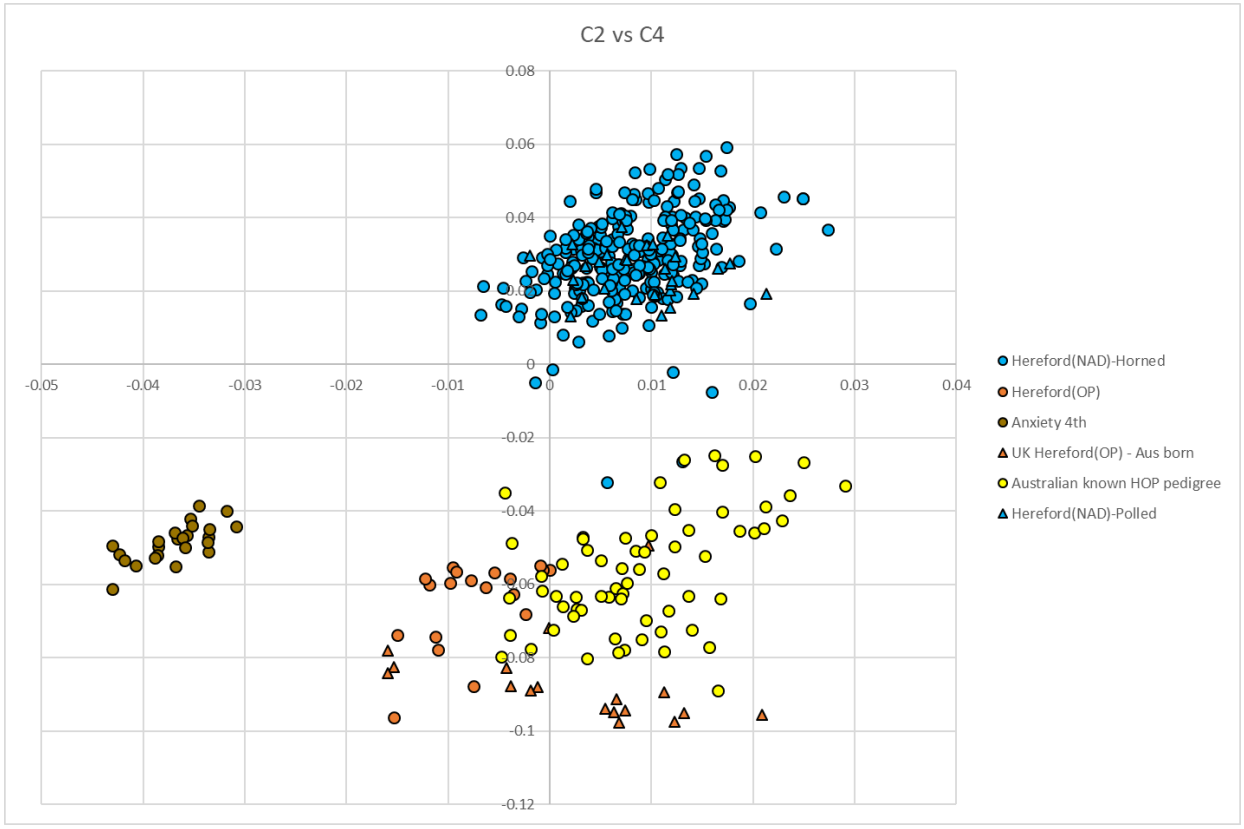
We plotted Hereford (OP) animals used by Ogden and Hereford (NAD) animals supplied by HHBI, the latter being divided into horned and polled subtypes. Also displayed are Australian animals with "E" registry numbers (referencing English origin and thus descent from the original UK Hereford (OP)), plus the Jim Lents' herd in the USA, which are derived from the Anxiety 4th bloodline, a prominent Hereford (OP) sire widely used in historical US breeding but no longer represented in the UK (barring re-importation). Finally, we include Australian-born Hereford (OP) animals imported as embryos from the UK.

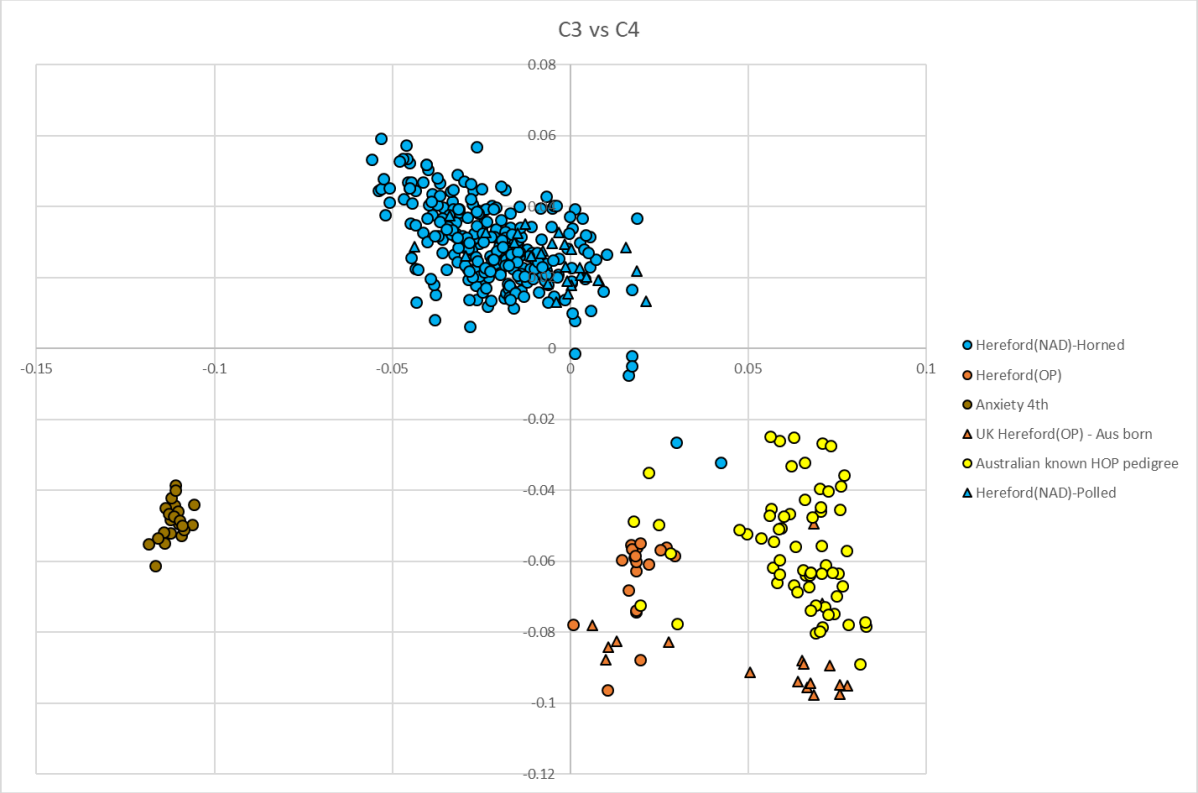



C1 vs C3
0.1
0.05
0
-0.05
-0.1
-0.15
0
0.02
0.04
0.1
0.12
0.14
0.16
0.18
Hereford(NAD)-Horned
Hereford(OP)
Anxiety 4th
UK Hereford(OP) - Aus born
Australian known HOP pedigree
Hereford(NAD)-Polled

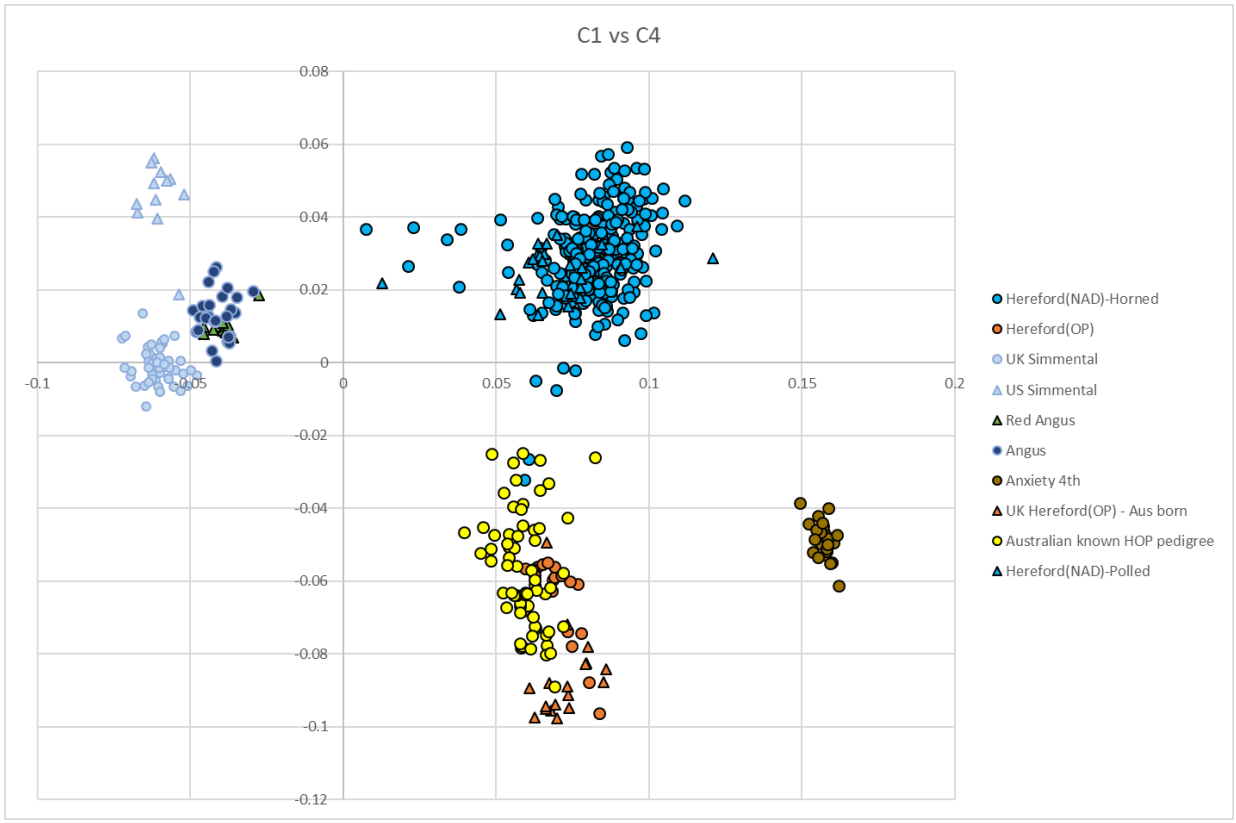

C1 vs C4
0.08
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
-0.08
-0.1
-0.12
0
0.02
0.04
0.08
0.1
0.12
0.14
0.16
0.18
Hereford(NAD)-Horned
Hereford(OP)
Anxiety 4th
UK Hereford(OP) - Aus born
Australian known HOP pedigree
Hereford(NAD)-Polled

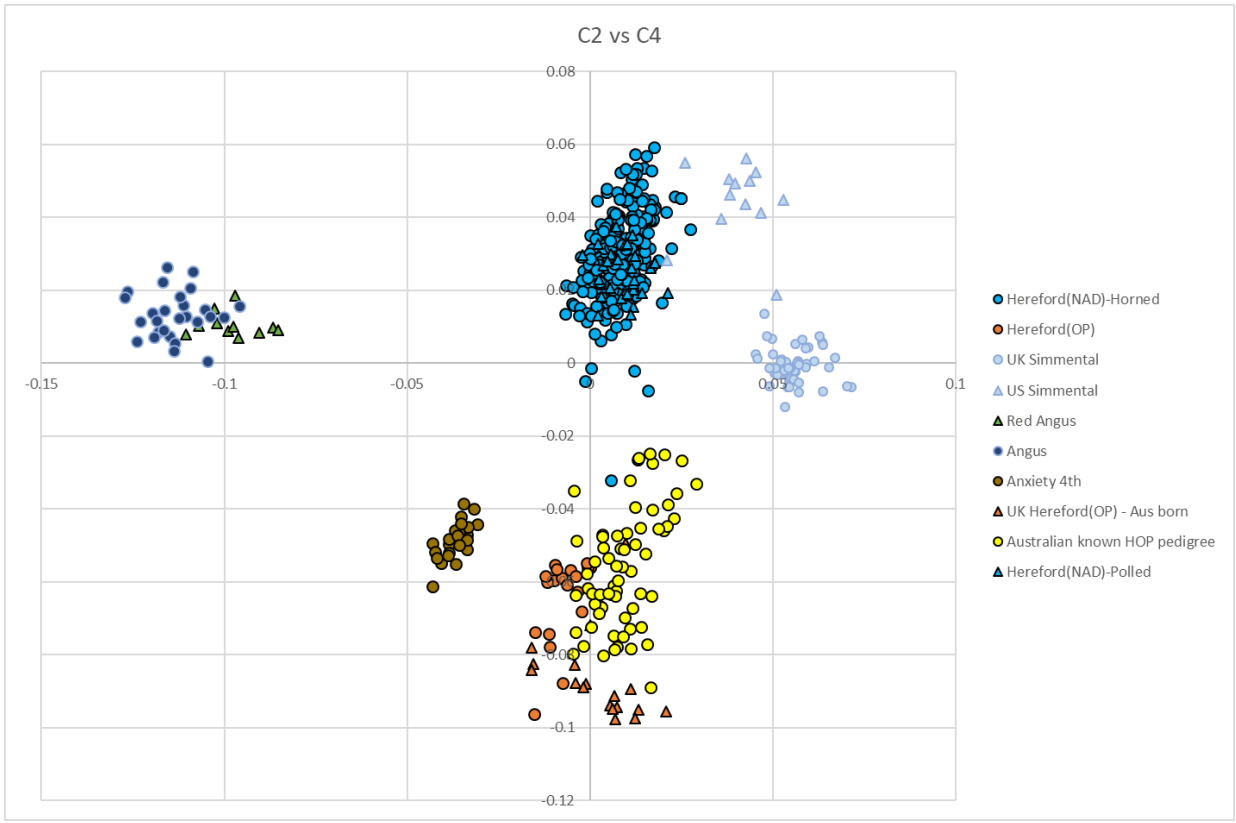

C2 vs C3
0.1
0.05
0
-0.05
-0.1
-0.15
-0.05
-0.04
-0.03
-0.02
-0.01
0
0.01
0.02
0.03
0.04
Hereford(NAD)-Horned
Hereford(OP)
Anxiety 4th
UK Hereford(OP) - Aus born
Australian known HOP pedigree
Hereford(NAD)-Polled


C2 vs C4
0.08
0.06
0.04
0
-0.02
-0.04
-0.06
-0.08
-0.1
-0.12
-0.05
-0.04
-0.03
-0.02
-0.01
0
0.01
0.02
0.03
0.04
Hereford(NAD)-Horned
Hereford(OP)
Anxiety 4th
UK Hereford(OP) - Aus born
Australian known HOP pedigree
Hereford(NAD)-Polled



Whilst viewing different clusters does alter the orientation or relative distances on the plot, the division between Hereford (NAD) and Hereford (OP) is clearly visible in all six cases, but most apparent in C1vsC4 and C2vC4. Consequently, it was decided to use these views for subsequent assessment of additional genotypes supplied for the project. To show the strength of the division between groupings, we also plotted two non-Hereford breed groups - UK and US Simmental profiles (as examples of genetic variation due to environment/breeding practice) and Angus/Red Angus (as examples of two recently diverged breeds).





The clear pattern of division between Hereford (OP) and Hereford (NAD) clearly suggests that the two have been separated by a significant genetic shift, not likely to result from genetic drift or different breeding practices.

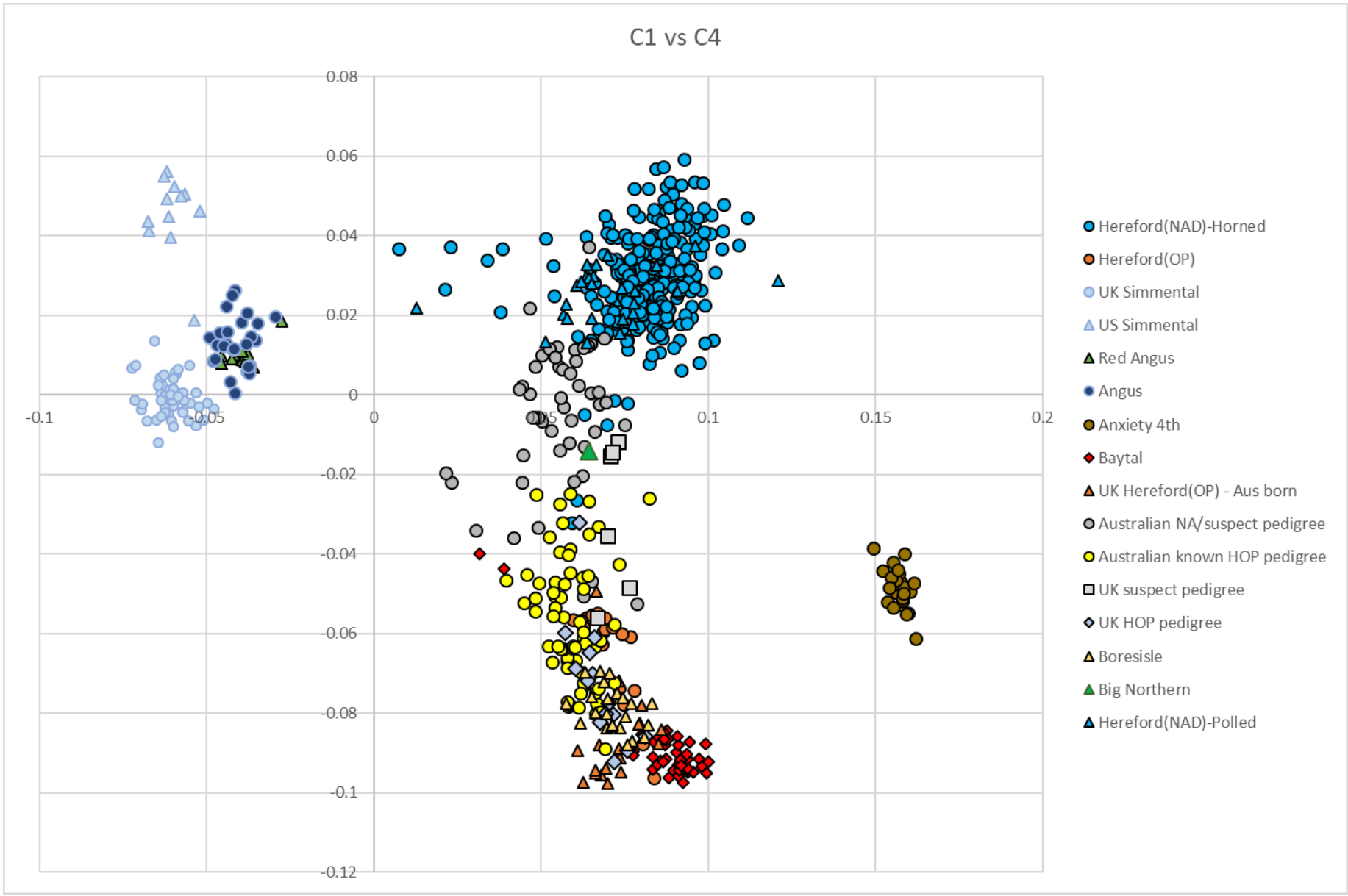
2.2. Classification of animals to Hereford (OP) or Hereford (NAD)

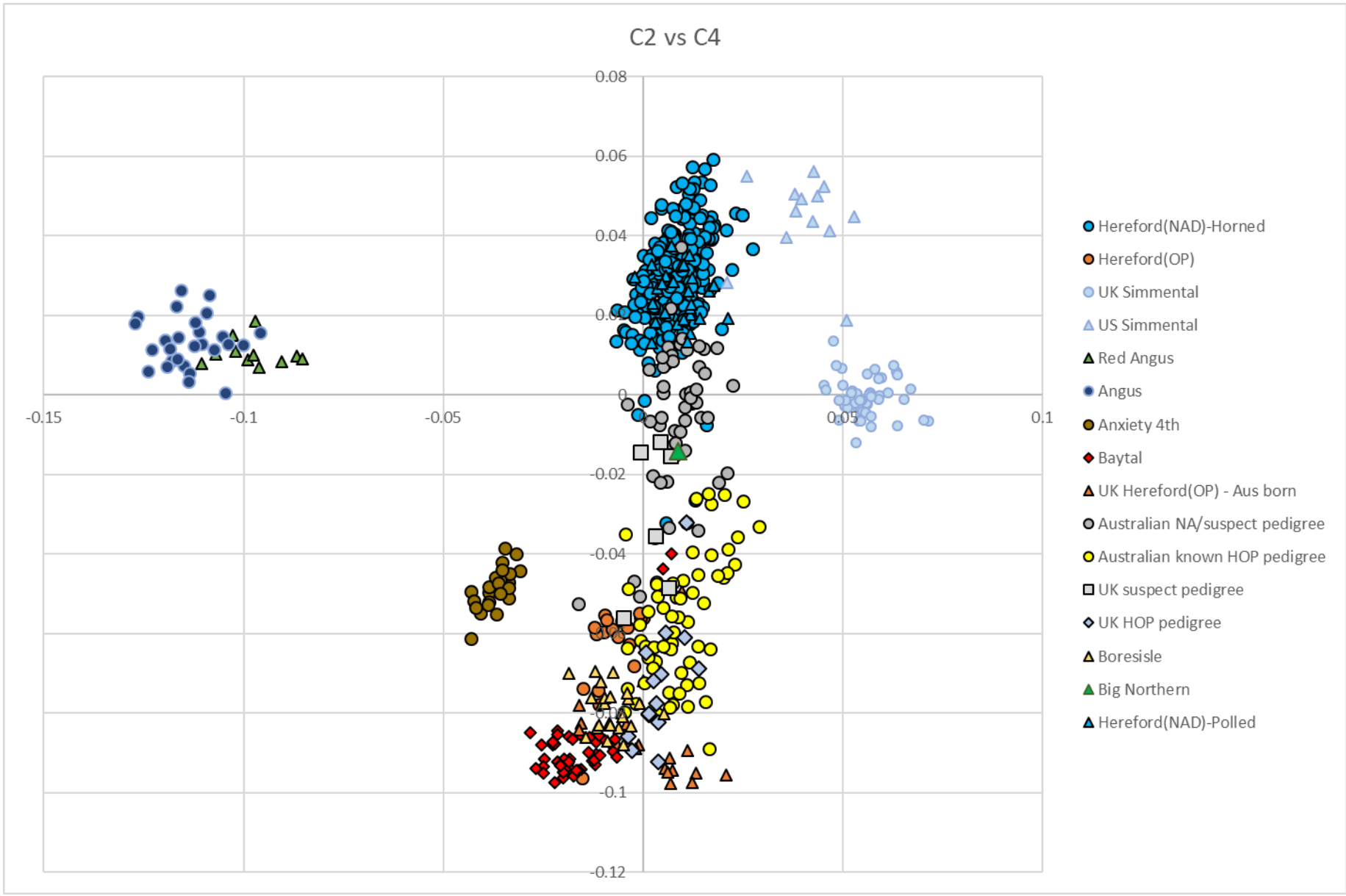
As demonstrated above and confirming the results of Taylor and Ogden, there is a strong genetic division between the Hereford (OP) and Hereford (NAD), which is unlikely to be due to differential breeding or environment. Rather, it is highly likely that the Hereford (NAD) has undergone hybridisation with genetic material from an outside source at some point in its history within North America and Australia, resulting in a composite that has subsequently been selected upon. If this is the case, then there is a strong argument for a registry test to maintain the Hereford (OP) as a distinct genetic resource, treating the two as separate breeds. Registering the two under a single Herd Book runs the risk of introgressing non-Hereford genetics into the Hereford (OP). Indeed, given the widespread use of North American/Australian-derived animals, this has already occurred.

To determine whether we could identify a cut off point for identifying animals as Hereford (OP) or Hereford (NAD), we next plotted a number of UK and Australian animals of known pedigree supplied by HHBI. In addition to the animals shown on the previous plots, these contained:

- UK animals with suspect Hereford (OP) pedigree
- UK animals of strong Hereford (OP) pedigree
- Australian animals with suspect pedigrees or known strong influence from North American bloodlines.
- We also included 48 animals from the UK Baytal and 24 individuals from the Boresisle Hereford (OP) herds.

Finally, for reference, we include Big Northern, a 1968 bull known from pedigree to represent a 50/50 mix of UK HOP and North American-derived bloodlines:

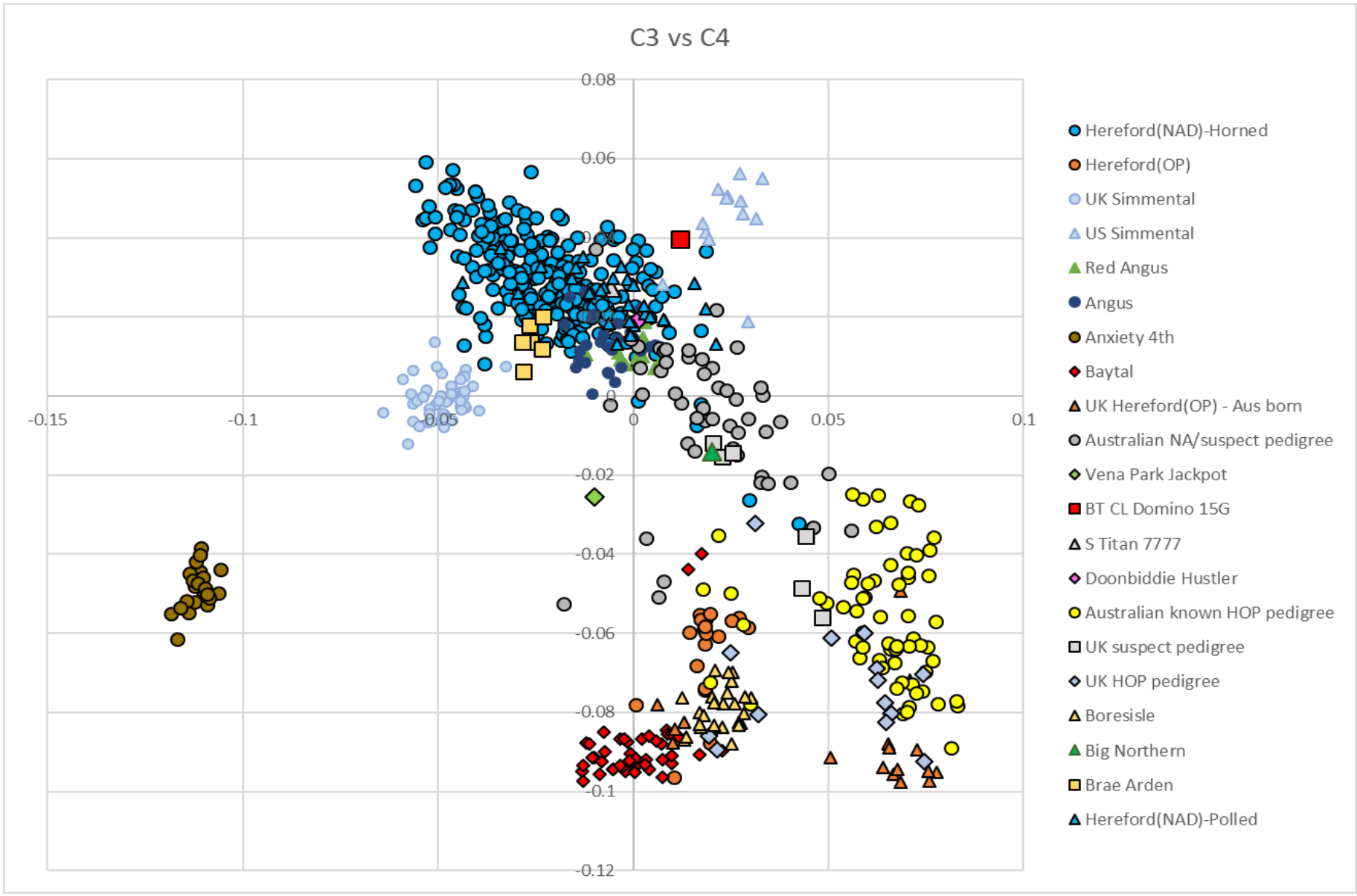

C1 vs C4
0.08
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
-0.08
-0.1
-0.12
-0.1
-0.05
0
0.05
0.1
0.15
0.2
Hereford(NAD)-Horned
Hereford(OP)
UK Simmental
US Simmental
Red Angus
Angus
Anxiety 4th
Baytal
UK Hereford(OP) - Aus born
Australian NA/suspect pedigree
Australian known HOP pedigree
UK suspect pedigree
UK HOP pedigree
Boresisle
Big Northern
Hereford(NAD)-Polled


C2 vs C4
0.08
0.06
0.04
0
-0.02
-0.04
-0.1
-0.12
-0.15
-0.1
-0.05
0.05
0.1
Hereford(NAD)-Horned
Hereford(OP)
UK Simmental
US Simmental
Red Angus
Angus
Anxiety 4th
Baytal
UK Hereford(OP) - Aus born
Australian NA/suspect pedigree
Australian known HOP pedigree
UK suspect pedigree
UK HOP pedigree
Boresisle
Big Northern
Hereford(NAD)-Polled

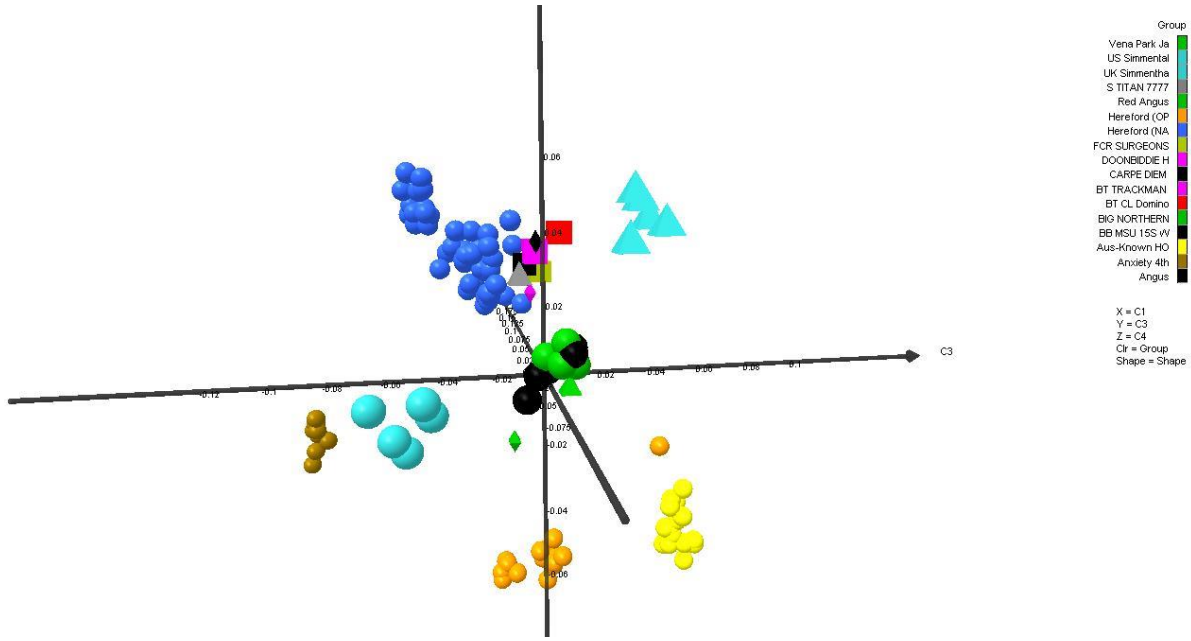
With only a few exceptions, the Hereford (OP) reference animals, together with animals of known strong UK Hereford (OP) pedigree almost all sit below the -0.02 line on the Y-axis (C4), whilst those with known Hereford (NAD) influence or suspect pedigrees lie above it. Big Northern (green triangle) behaves exactly as predicted based on strong pedigree records, lying halfway between the two reference groups. The Baytal herd lies slightly off from the rest of the UK reference Hereford (OP) samples as a whole, but this herd is line-bred and this may account for the shift. It should be noted that two of the animals from the Baytal herd plotted here were also represented in the Hereford (OP) reference set data drawn from that used by Ogden in his FAO112 report and the replicate genotypes match well, demonstrating that the test is consistent with earlier work.

This then presents the basis for a “Yes/No” assignment test if the decision is taken to treat the Hereford (OP) and the Hereford (NAD) as two separate genetic groups. It also demonstrates that some Australian individuals may already be have been registered as Hereford (OP) when they carry a significant amount of North American-derived or other external genetics. However, we might have issues developing the test beyond “Hereford (OP)/Hereford (NAD)” – i.e. if we wish to determine levels of admixture in animals that straddle the line. However, the inclusion of Jim Lents’ Anxiety 4th-derived herd gives more insight.

The Lents’ animals lie slightly below the cut off on the Y-axis but are much more extreme on the X-axis (C1 or C2), though moreso on C1. This herd is highly linebred, as can be seen from the tight clustering, and used dam lines that no longer exist within the UK Hereford (OP). A similar effect, though much less extreme, can be seen in the Baytal herd. Anxiety 4th was highly used in US bloodlines: earlier plots had shown that C3 appeared to demonstrate some relationship between the Hereford (NAD) and the Anxiety 4th-based Lent’s herd. To see if there is evidence of this influence, we examined the plot for C1vsC3 and plotted several animals from the Brae Arden herd, which are known to have strong Anxiety 4th influence but also derive from other extinct UK bloodlines. For comparison, we also show BT CL Domino 15G (polled) and S Titan 7777 (horned), two Hereford (NAD) animals known to have had strong influence on the UK and US populations, respectively. Similarly, we also plotted Doonbiddie Hustler (polled), an Australian sample that has contributed heavily to UK modern Hereford genetics.


C3 vs C4
0.08
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
-0.08
-0.1
-0.12
-0.15
-0.1
-0.05
0
0.05
0.1
Hereford(NAD)-Horned
Hereford(OP)
UK Simmental
US Simmental
Red Angus
Angus
Anxiety 4th
Baytal
UK Hereford(OP) - Aus born
Australian NA/suspect pedigree
Vena Park Jackpot
BT CL Domino 15G
S Titan 7777
Doonbiddie Hustler
Australian known HOP pedigree
UK suspect pedigree
UK HOP pedigree
Boresisle
Big Northern
Brae Arden
Hereford(NAD)-Polled

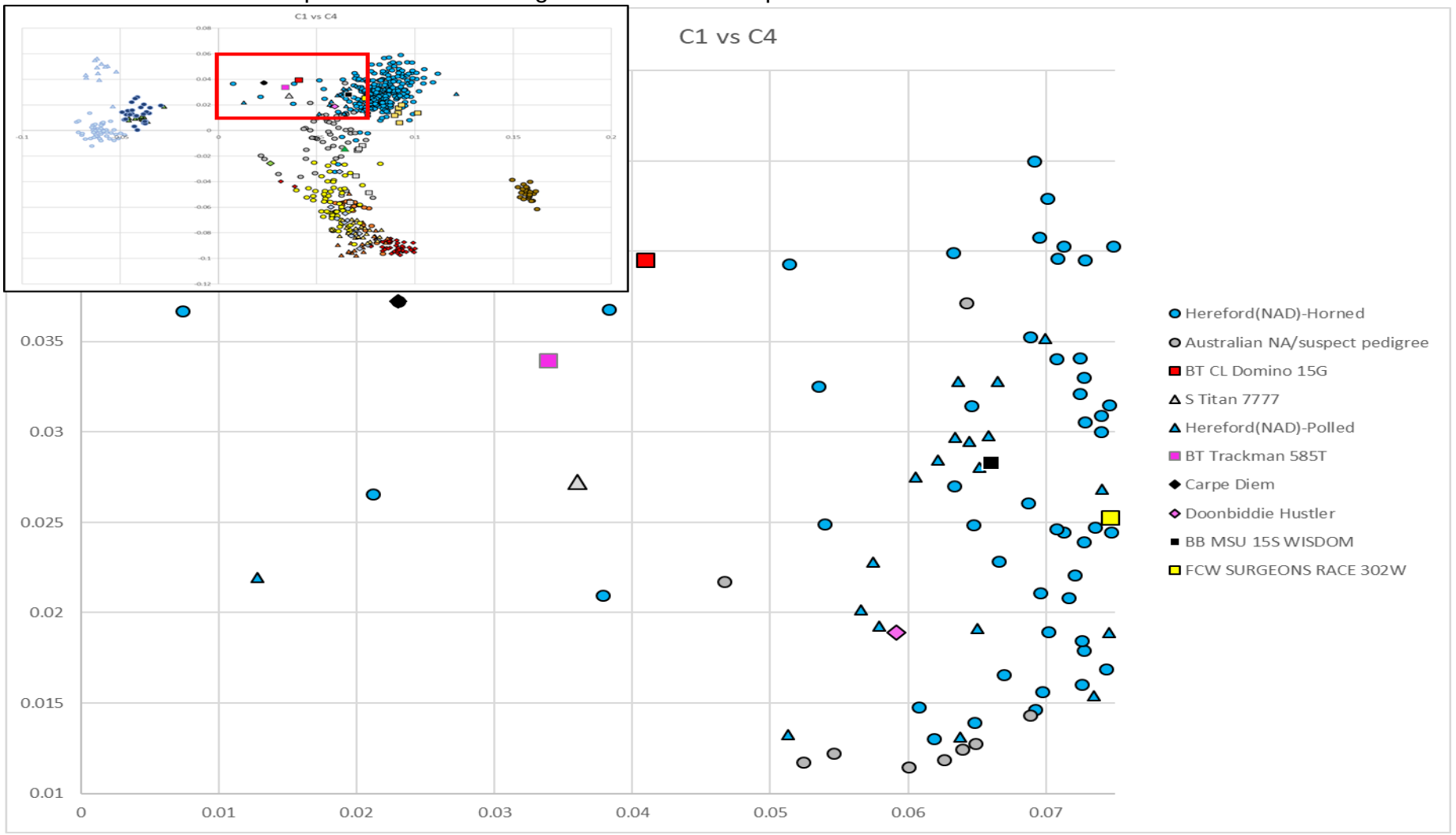
As can be seen, there is some movement on this plot within the Brae Arden individuals towards the Anxiety 4th, though this is weak. This suggests that we can detect historical bloodlines within the Hereford (NAD) and that, whilst the majority of Hereford (NAD) are clearly not aligned to the UK and Australian-Hereford (OP) pedigree animals, they could be judged relative to Anxiety 4th and other key historical pedigree lines (i.e. Line 1). This is investigated further (section 2.3) via population structure testing using the Admixture software package (Alexander *et al.*, 2009). Note that the apparent clustering of Angus with Hereford (NAD) in this plot should not be taken as indicative – this is a limitation of viewing multidimensional data in a 2-D plot format: if the plot were viewed in a 3-D format, the Angus clusters lie substantially "above" the Hereford (NAD). To demonstrate this, we used the commercial Excel plug-in software XLSTAT (https://www.xlstat.com/en/) to create a 3-D plot of C1vsC3vsC4 showing a reduced selection of animals from each of the main groupings for easier viewing (including some named animals mentioned below). A screengrab of the plot is included here but is again limited by representation as a 2-D image.



For a fuller view of the data, we have uploaded a video file of the plot being rotated (with explanatory voiceover) to the following link: https://youtu.be/R4VQjkB48Jc

How much can these plots tell us when determining corrupt genetics? To check this, we plotted the C1vsC4 locations for offspring of two individual bulls known to be corrupt: Perfection (deregistered by the American Hereford Association in 1986) and Sweet Coed Anxiety (a Canadian cow whose owner was stripped of membership in 1989). Both animals had been widely used and, despite being deregistered, descendants are still present. We used BT Trackman 585T (itself suspected of hybrid genetics) for Perfection ancestry (son) and BBU MSU 15S Wisdom and FCR Surgeons Race 302W for Sweet Coed Anxiety (great great grandchildren). We also include Carpe Diem – this animal is also registered with the American Chianina Association but has ~50% Hereford (NAD) genetics by pedigree – the AHA have

assigned it a “T” registry number in an effort to prevent it being used in Hereford breeding. To make the samples easier to see, we zoom in on the area of the plot in the red rectangle shown in the subplot:



It can be seen that descendants of known corrupt individuals and known/suspected hybrids are positioned well to the left of the C1vsC4 plot within the highlighted area. Perhaps worryingly for breeders, so are S Titan 7777 and BT CL Domino 15G, two widely used sires in the modern US and UK herds, respectively. To highlight the potential issues if animals in this plot are problematic: from the 78 animals to the left of (and including) the known corrupt Sweet Coed Anxiety descendant FCR Surgeons Race 302W, there are 62,772 registered progeny in the US and Australia alone (they have also been used in South America, UK and Canada). As some of these animals date back to the 1960s (see below), the number of additional generation descendants may be over a million:

Decade	**Number of animals in plot**	**Number of progeny**
1960s	2	275
1970s	13	15,812
1980s	17	19,991
1990s	11	3,825
2000s	17	13,552
2010s	15	9,317

However, it is worth stating that positioning of animals within this region of the plot is not, in itself, sufficient to state that the animal has corrupt genetics – just that the animal must be investigated further using the tests we subsequently discuss in sections 2.3 and 2.4 and viewed in the context of all three metrics before making a decision.

Also of interest from this plot is that, of our 27 other polled US samples, 23 are contained within the red highlighted region – most of which have progeny in the UK. The polled trait is dominant: breeding a polled animal to a horned animal always produces polled offspring if the polled parent carries both copies of the polled gene and has a 50/50 chance of polled offspring if the polled parent only carries one copy (these offspring will then also only have one copy). The origin of the poll in the Hereford is contended. An account given in MacDonald and Sinclair (1909) describes the “accidental” mating of a Hereford cow with a Red Polled bull by a breeder in Iowa in 1894, resulting in a polled calf which the breeder allowed to mate with predominantly Hereford crossbred cows. Around 40% of the bulls and 75% of the heifers from this population were polled. He then bred polled heifers containing the most Hereford blood with bulls from another breeder from Kansas – Guthrie - who had used a similar practice (only with Polled Durham – nowadays referred to as Polled Shorthorn) to create a polled population “a few years earlier”. The two together selectively bred polled animals that were fixed (homozygous) for the polled trait.

A later account, however, by Ornduff (1957) details that Guthrie and other breeders following similar practices were not able to register their animals or their descendants with the American Hereford Record, as the standard excluded non-purebreds. Ornduff records that a breeder named Warren Gammon, who founded the American Polled

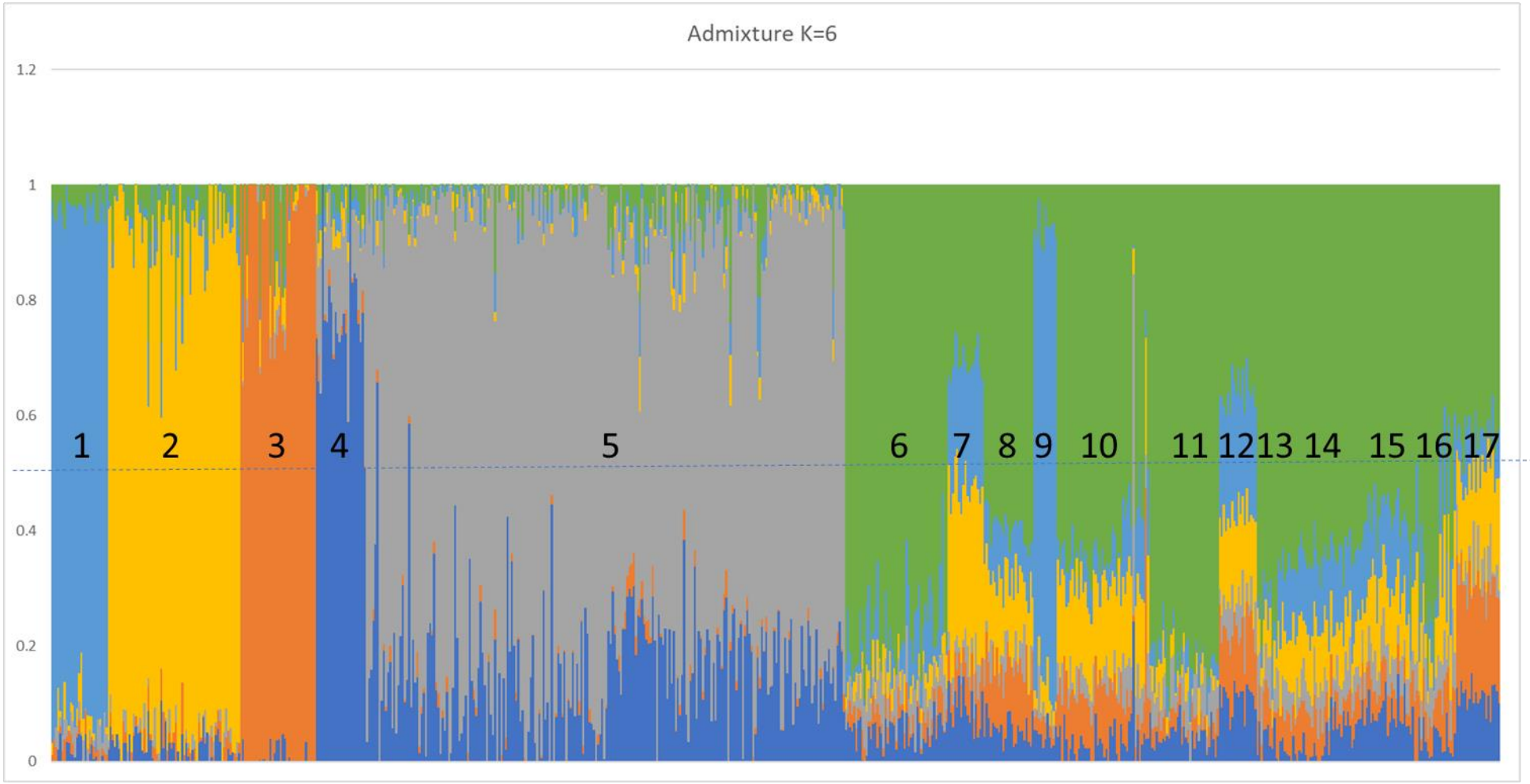
Hereford Cattle Club in 1900, had enthused about the work of Guthrie *et al.* but foresaw these issues and committed to finding "freak" naturally polled Herefords from which to create a "double standard" polled Hereford. Ornduff states that Gammon put out a call to members of the American Hereford Cattle Breeder's Association, which ran the American Hereford Record, asking for examples of polled animals. Fourteen (4 bulls and 10 cows) were identified and Gammon bought all 4 bulls and 7 cows for use in his new herd, later acquiring two more for a founder population of 13 animals. Ornduff does not record how or where these animals were found, nor how Gammon confirmed their purebred status beyond the fact that the AHR had accepted them. Ornduff does note, however, that the American polled Hereford was later used to introduce the trait into the UK, where it did not previously exist. Genetically speaking, it strains credibility that the US population – exported from the UK originally – happened to have 16 "freak" polled Herefords existing simultaneously with which to form the genetic base of a new population when this trait had never been observed in the UK in records going back to 1797 (Heath-Agnew, 1983). Given the dominant nature of the polled trait, any appearance of a polled bull would have a 50% chance of producing offspring, so barring aggressive selection against mating such individuals (which would require strict recordkeeping in any event), the absence of polled in the UK record makes it improbable that the trait existed cryptically and was only noticed in the US following export. The other possibility – a random mutation in the US that created the trait *de novo* – is even more unlikely. The polled locus in European cattle is the same in all breeds (there is a separate polled mutation in Zebu cattle) and thus almost certainly predates the original division of those breeds from a common ancestral type. It is more probable that what Gammon identified as "purebred" polled Herefords were the result of unrecorded hybridisation (deliberate or accidental) with polled breeds that had been bred back to Hereford as per the practices of Guthrie and others.

The close clustering of the polled animals is intriguing, not just for their positioning within the area of suspect genetics shown by Perfection and Sweet Coed Anxiety descendants, but also because it further suggests (as did the Brae Arden samples) that we might be able to identify bloodlines within the North American samples. To do this, and to further assess our ability to look at percentages of Hereford (OP)/Hereford (NAD) genetics within samples, we turn to admixture testing.

2.3. Admixture testing for population assignment

Briefly, this method tests the assumption that the overall data can be split into two or more groups (*K*) of samples based on their genetic relatedness – run in "unsupervised mode", the software has no prior knowledge of which animals come from which breed/population. Once groupings are identified, each animal is given a percentage assignment score to that grouping. Each grouping is then manually checked to see which animals fall within it and thus whether a grouping is associated with animals of a particular breed. Besides the groups used in the plot above, other breeds included in this analysis were Angus, Red Angus, Holstein, Jersey, Guernsey, Simmental,

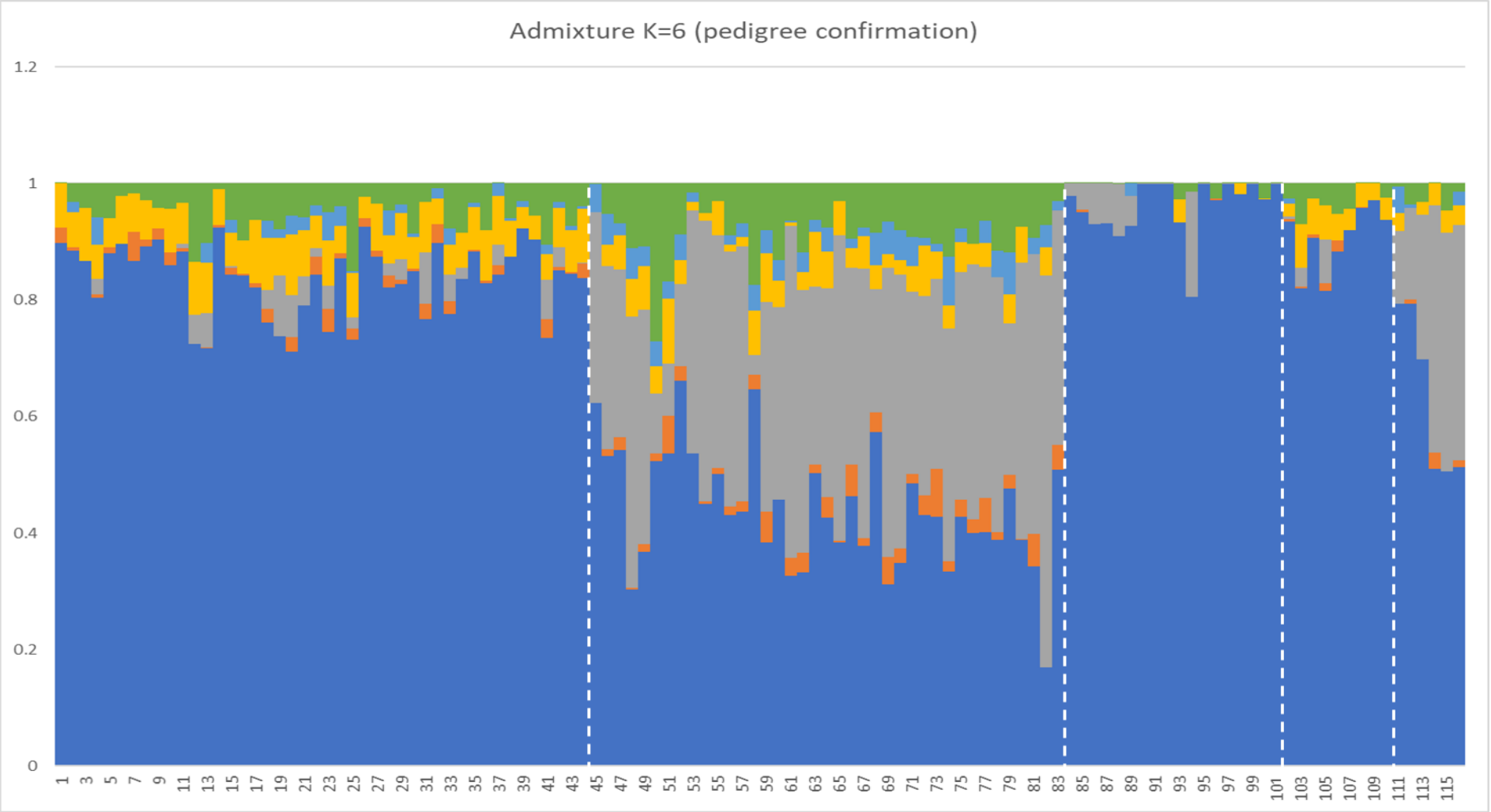
Limousin, Romagnola, South Devon, Brown Swiss, British Blond, Charolais, Norway Red and Piedmontese, plus the assorted crossbreeds from the carcass verification work. The Anxiety 4th animals were found to rapidly form their own unique grouping (which never showed admixture with other animals) even at low *K* values so were excluded from this analysis. The relationship between the Hereford (NAD) and Hereford (OP) Herefords is apparent at *K*=4, where manual examination of the animals falling into each group shows the groupings to be Angus, Holstein, Hereford and "Other". However, the clear distinction between the Hereford (NAD) and the Hereford (OP) is born out here even at low *K* values – at *K*=6, the software identifies the division between the two groups (Jersey cattle forming their own grouping at *K*=5). Shown below are the known breeds to demonstrate the clear divisions:



The major groupings are 1) Angus, 2) Holstein, 3) Jersey, 4) Hereford (OP) and 5) Hereford (NAD). Note that the final sample in the "Hereford (OP)" group is Big Northern, which as expected from both pedigree and our cluster analysis comes up as 50% Hereford (OP)/50% Hereford (NAD) and is shown by the demarcation line. As would be expected due to their close relationship, the Red Angus

(9) shows a high assignment to the same light blue grouping as the black Angus. The “Other” grouping contains various amounts of assignment for 6) Limousin, 7) Norway Red, 8) Piedmontese, 10) Romagnola, 11) Simmental (note the odd “peak” samples separating this and Romagnola are the US Simmentals – the unusual behaviour of these is discussed later), 12) South Devon, 13) British Blonde, 14) Brown Swiss, 15) Charolais, 16) Friesian and 17) Guernsey. Of these, the Limousin and the Simmental have the highest assignment to this grouping.

Whilst, at this *K* value, the Hereford (NAD) shows some admixture with the Hereford (OP) (as would be expected), the division between the two is obvious and it is striking that the split between the two occurs before the software can distinguish separate breeds such as the Limousin and the Simmental. The ability of the software to provide percentage assignments to each grouping is useful as we can now look at the Australian “known Hereford (OP) pedigree” and “NAD/Suspect Pedigree” animals in more detail. Also included are the “UK Hereford (OP)-Aus born”, “UK-Hereford (OP) pedigree” and “UK-suspect pedigree” individuals for comparison:

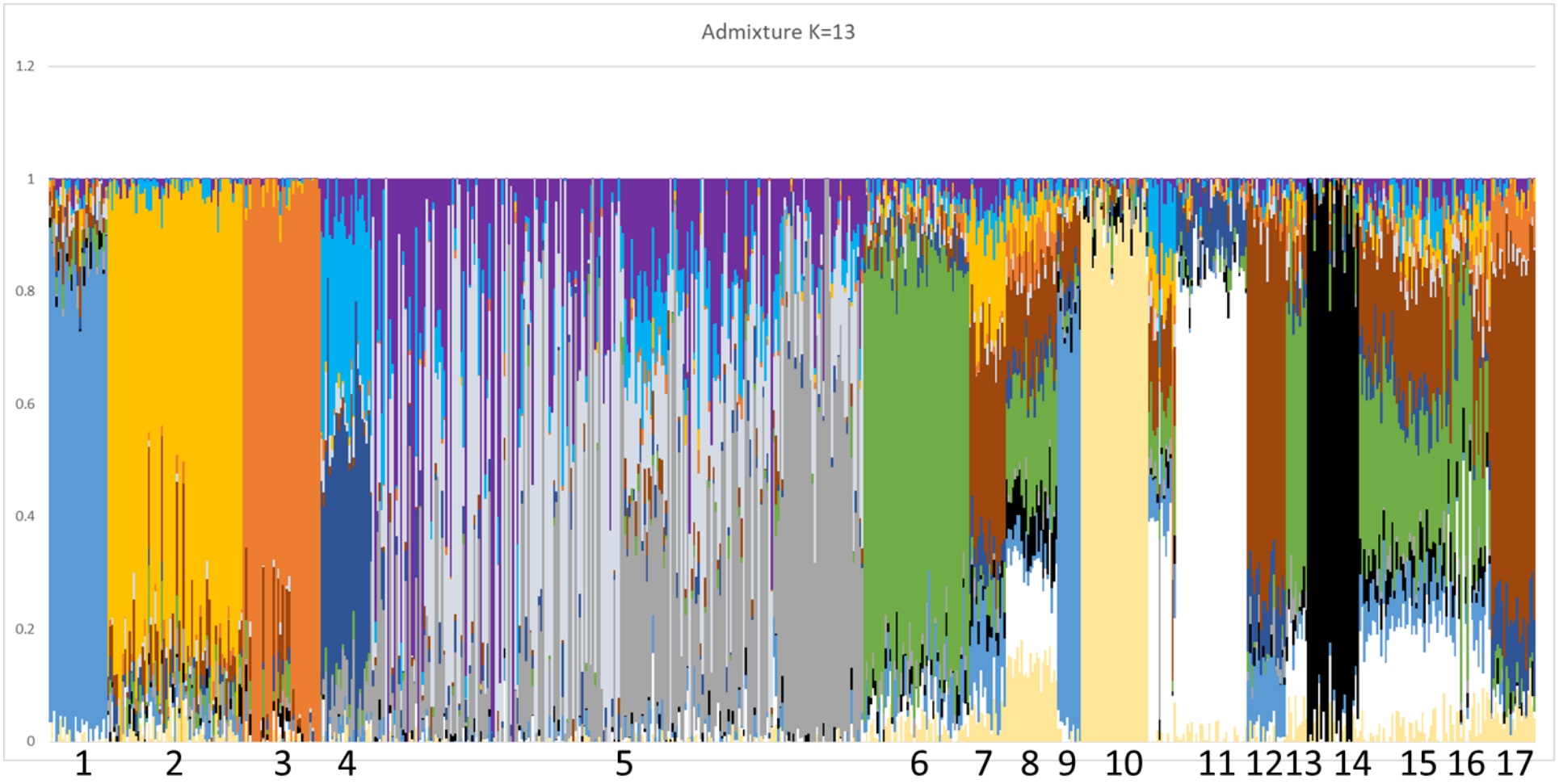


As before, dark blue represents the Hereford (OP) grouping and grey the Hereford (NAD). Samples 1-44 are Australian-known Hereford (OP) pedigree, 45-83 are Australian-Hereford (NAD)/suspect pedigree, 84-101 are UK Hereford (OP)-Australian born, 102-110 are UK-Hereford (OP) pedigree and 111-116 are UK-suspect pedigree.

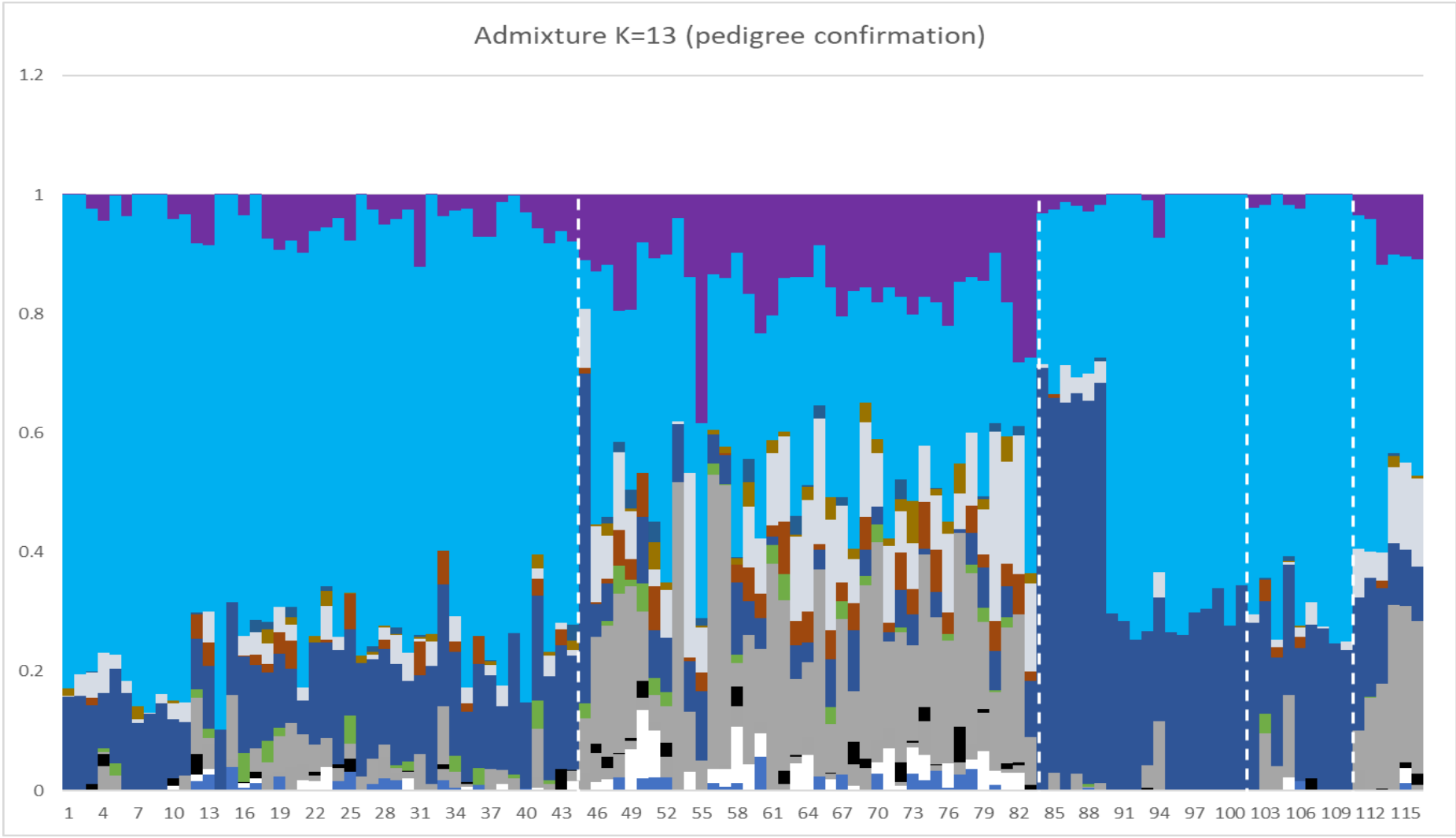
These results confirm the positioning of the animals on the cluster plots, but now give us clear metrics on percentage of NA genetics within each animal.

As the basis for a Hereford (OP)/Hereford (NAD) test, this is straightforward and further demonstrates the clear distinction between the two groups. Taking it further, however, we confirm our suspicions from the cluster analysis that we can identify different groupings within the Hereford (NAD) and Hereford (OP) samples. At *K=7*, we began to observe a division of the Hereford (OP) into two groups, one primarily observed in Australian samples, and one predominately grouping some of the UK reference set animals as well as the Baytal and Boresisle individuals. This split, then, would appear to derive from differences in the maternal Hereford (OP) bloodlines still extant in the UK and Australia, which only partially overlap. We thus tentatively term these groupings HOP-UK and HOP-Australia, although these are broad trends and some UK animals may have relatively high assignment to HOP-Australia and *vice-versa*.

By *K*=10 (*K*=8 and *K*=9 had split off a Romagnola group and divided the rest into two subgroups, one predominated by Limousin and the other by Brown Swiss), the software had split the North American samples into two groups. One of these was predominated by animals with significant amounts of US-Line 1 breeding, the other a mixture. At *K*=11, these two groups were re-merged due to the split of the Simmental individuals from the “Limousin” and “Other” categories, but split apart again at *K=12* and further divided at *K=13* into three historical groupings apparent from the animal pedigrees. These were US-Line 1, US-Mark Donald and Canadian-influence bloodlines. This plot is shown below – order is the same as the K=6 plot:



This has potential for showing the ancestry of Herefords (NAD), but is also useful when looking at Australian and UK samples with Hereford (OP) influence or known NA/suspect pedigree:



As before, blocks are in order: Australian-known Hereford (OP) pedigree, Australian-NA/suspect pedigree, UK-Australian born, UK-Hereford (OP) pedigree and UK-suspect pedigree. The dark blue is the group predominated by the HOP-UK grouping with the light

blue being the HOP-Australia grouping. Purple is the US-Mark Donald bloodline, dark grey Canadian influence and light gray US-Line 1.

Again, the suspect Australian animals, or those with known Hereford (NAD) influence, behave as expected and show a lack of the current UK grouping seen in the Australian-born UK animals and the Australian-known Hereford (OP) pedigree, neither of which show significant levels of Hereford (NAD) ancestry (it should be noted that cases where small degrees of assignment (<10%) are made to NAD groups most likely represent noise due to the shared history of the Hereford (OP) and the Hereford (NAD), rather than actual NAD influence). We can thus begin to put hard numbers on animals straddling the line on the cluster plots: for example, Big Northern (known to be 50% NAD/50% HOP by ancestry), gives a result at *K*=13 of:

Canadian influence: 26.8%

HOP-UK: 11.7%

US-Line 1: 10%

HOP-Australia: 33.1%

US-Mark Donald: 12.5%

With the remainder being slight mis-assignment to other breeds. Summing the HOP types gives us 44.8% and the US types 49.3%.

Meanwhile, a similarly positioned animal on the cluster plots – Vena Park Jackpot – shows only a combined 24.8% NAD influence and 41.2% assignment to HOP (primarily the HOP-UK grouping at 33.6%). The remainder is assigned to other breeds, with Simmental highest at 10.1% assignment and Limousin at 7.4%, equal to the proportion of this animal's genetics assigned to any of the three NAD groups. This animal appears to sit slightly closer to HOP on the cluster plots than Big Northern, despite its lower overall assignment to HOP, but at only 65.9% assignment to either "Hereford" grouping it could be considered a benchmark for rejecting animals. In fact, the positioning of this animal on the C1C4 plot is somewhat misleading due to the issues caused by representing the data as 2-dimensional plots, as discussed in Section 2.2. In fact, when viewed using a 3-D plot (use link provided in Section 2.2), it can be seen that, as well as sitting well to the left of Big Northern, it also lies well "behind" it. This unusual positioning is actually consistent with the fact that there is no recorded North American influence in this animal's pedigree, suggesting that the 24.8% assignment to NAD is actually a "best guess" on behalf of the software.

The known corrupt NA individual discussed earlier, BT Trackman 585T, similarly only has 60.5% assignment to NA and 11% assignment to HOP, for a combined 71.5%.

The bulk of the remainder is a 12.5% assignment to Simmental… this animal has long been suspected of hybrid genetics – a 12.5% assignment could indicate a 2nd generation backcross to Hereford (NAD).

Note that this combination of Hereford (NAD) and Hereford (OP) admixture values may be useful for identifying individuals with more recent suspect breeding or large introgressions of foreign genetics, but does not reflect that the modern Hereford (NAD) is likely to have had introgressions as well. As discussed in section 2.4, the current assay provides the basis for a test to divide Hereford (OP) and Hereford (NAD) – determining what foreign influence has occurred in the Hereford (NAD) historically is more complex and will require further experiments.

Higher values of *K* split out some of the other breeds (Guernsey, South Devon and Brown Swiss), but there is no further subdivision of the Hereford (OP) or Hereford (NAD) types until *K*=17, where the Brae Arden samples form their own grouping. By this point, a group of samples from our carcass project had also formed – back checking these showed that they were Aberdeen Angus but all came from the same closed herd. As discussed in McMahon *et al.* (2015), we saw some potential for genetic testing to identify farm (or at least region) of origin in highly stratified genetic systems such as linebred cattle. This may be of interest to both Hereford (OP) and Hereford (NAD) breeders in future as a component of animal traceability.

2.4. Identification of markers strongly differentiating Hereford (OP) and Hereford (NAD)

To determine which markers within the set of 4593 SNPs contribute the most to differentiation between the two groups, the Hereford (OP) and Hereford (NAD) reference sets were assigned case/control status within PLINK (HOP set as case), with all other animals set to “unknown” status. An allelic case/control test was then performed to identify markers displaying significant differences in allele frequency between the two groups. This identified 63 markers of genome-wide statistical significance ($p < 1E\text{-}07$) following Bonferroni multiple testing correction. Analysis of the genomic locations of these showed that the majority of these are in/near genes involved in milk yield/composition or body mass, as would be expected given the phenotypic differences between the two groups.

CHR	SNP	BP	Freq_HOP	FREQ_NA	P	Genes nearby
0	ARS-BFGL-NGS-5133	3553169	0.1552	0.6522	2.00E-07	Near CSMD1 - wither height link in buffalo and myoclonic epilepsy in humans
0	ARS-BFGL-NGS-96496	6388642	0.08621	0.5227	9.95E-07	In GPM6A - behavioural? Linked to educational attainment/schizophrenia.
0	ARS-BFGL-NGS-88529	9697693	0.03448	0.5	3.43E-08	Near RYR2 - associated with somatic cell count/mastitis
0	ARS-BFGL-NGS-103387	35036603	0.1034	0.6304	1.62E-08	In ZMIZ1 - strong candidate marker for female fertility in Holstein study
1	Hapmap60044-rs29010295	20634675	0.05172	0.5652	6.64E-09	Near USP25 - linked to milk yield in buffalo study
1	BTB-01965833	34179096	0.01724	0.3913	9.44E-07	Near CADM2 - linked to body weight in cattle
1	ARS-BFGL-NGS-18743	53630189	0.1724	0.6739	2.02E-07	In MYH15 - associated with pulmonary hypertension "brisket disease", QTL for fatness in pigs
1	BTB-01568926	1.1E+08	0.1379	0.6957	6.38E-09	Near SHOX1 - linked to short stature in humans
1	ARS-BFGL-NGS-103480	1.56E+08	0.01724	0.5217	2.23E-09	Near PLCL2 - linked to tenderness and marbling in Nellore cattle
2	BTB-01242342	42336766	0.01724	0.413	3.64E-07	In GALNT3 - linked to ADG and various beef composition traits
2	BTB-01651888	59760905	0.01724	0.4565	5.10E-08	In THSD7B - linked to milk yield and in region associated with age at 1st calving
3	ARS-BFGL-NGS-22796	1.1E+08	0.1552	0.6087	1.60E-06	Near MRPS15 - linked to C15 fatty acid in Holstein milk
4	Hapmap39425-BTA-70290	10737673	0.01724	0.6739	5.75E-13	Near CALCR - linked to milk fat percentage
4	ARS-BFGL-NGS-86056	82796269	0.01724	0.3913	9.44E-07	In VPS41 - link to viral resistance
4	BTB-01290330	87744523	0.06897	0.5435	8.34E-08	In CADPS2 - linked to eating behaviour and weight in pigs, tenderness in cattle
5	ARS-BFGL-NGS-6826	12725994	0.05172	0.5435	1.96E-08	Near TMTC2 - part of cattle growth trait network
5	Hapmap25585-BTA-150007	45420592	0	0.4348	2.30E-08	In QTL region for milk yield in goats
5	ARS-BFGL-NGS-98210	46454015	0.03448	0.4348	6.89E-07	Near DYRK2 - linked to udder development (whole region is strong QTL)
5	ARS-BFGL-NGS-45453	1.13E+08	0.03448	0.4348	6.89E-07	Near MCHR1 (melanin concentrating hormone receptor - linked to growth/feeding behaviour and estrous fertility traits in cattle)
6	Hapmap54432-rs29022442	12119048	0	0.413	6.16E-08	Near NDST4 - linked to diabetes (other genes in pathway linked to DMI in cattle)
6	Hapmap38696-BTA-78192	15978059	0.08621	0.6522	1.31E-09	In PITX2 - linked to several growth traits
6	Hapmap32447-BTC-033214	33685443	0.01786	0.3913	1.43E-06	In GRID2 - linked to behavioural adaptation in comparison of T3 (1960–1980 SDM-1965 vs. Holstein)
6	ARS-BFGL-NGS-86825	86794730	0.1897	0.6739	5.77E-07	Near UGT2A1 - linked to milk yield/composition
6	ARS-BFGL-NGS-82167	98952769	0.06897	0.4783	1.69E-06	Near ENOPH1 - region associated with rib eye area
6	Hapmap27089-BTC-045849	1.01E+08	0.03448	0.4783	9.56E-08	In WDFY3 - linked to teat placement in Holsteins
7	Hapmap49872-BTA-115580	89967934	0.05172	0.6522	6.39E-11	Near TMEM161B - in QTL region for stature in Holsteins
7	BTA-20841-no-rs	92201609	0.05172	0.5652	6.64E-09	Near CETN3 - linked to udder placement in US Holsteins
7	Hapmap60226-rs29011503	92744435	0.06897	0.5	6.37E-07	Close to above but in ADGRV1 - linked to tenderness
8	ARS-BFGL-NGS-5096	712643	0.1207	0.5652	1.32E-06	In PALLD - identified as ADG/DMI candidate in US Hereford
8	ARS-BFGL-NGS-112336	23409295	0.06897	0.5435	8.34E-08	In FOCAD - linked to meat traits inc marbling in Bonsmara and Pakistani cattle, also conception at 1st service
8	BTA-52455-no-rs	90808222	0.1379	0.587	1.49E-06	Near S1PR3 - associated with marbling and growth traits
8	BTB-01515859	1.09E+08	0.08621	0.587	3.90E-08	Near BRINP1 (possible link to fertility drop in high protein diets in cows? Also linked to age at puberty in pigs).
9	BTB-00394801	59383405	0.2069	0.6957	5.42E-07	Likely same region as below but also near EPHA7, involved in meat quality
9	Hapmap43073-BTA-83925	60786164	0.01724	0.6304	7.11E-12	Near BACH2, key regulator of milk FA profile. Also near MAP3K7 involved in bovine TB susceptibility
9	Hapmap47795-BTA-98172	64229989	0.03448	0.6304	4.14E-11	Same likely region as above but near HTR1E, involved in lactation
10	Hapmap48024-BTA-62291	37793494	0.08621	0.5217	8.69E-07	In GANC - linked to fatness and dry period in milk production
10	ARS-BFGL-NGS-69005	75821328	0.1552	0.6087	1.60E-06	In RHOJ, linked to endometriosis in humans
11	ARS-USMARC-Parent-DQ837643-rs29018818	66341589	0.1552	0.6304	5.76E-07	Near WDR92 - identified in comparison of two Italian breeds differing for tenderness
11	BTA-101061-no-rs	66450428	0.1034	0.6087	4.89E-08	Near WDR92 - identified in comparison of two Italian breeds differing for tenderness
12	ARS-BFGL-NGS-34948	53080458	0.1379	0.6957	6.38E-09	In SCEL - associated with dairy character in Chinese Holsteins
12	ARS-BFGL-NGS-16052	71469688	0.1296	0.6522	6.88E-08	Near DCT, linked to white spotting in Holsteins and Friesans

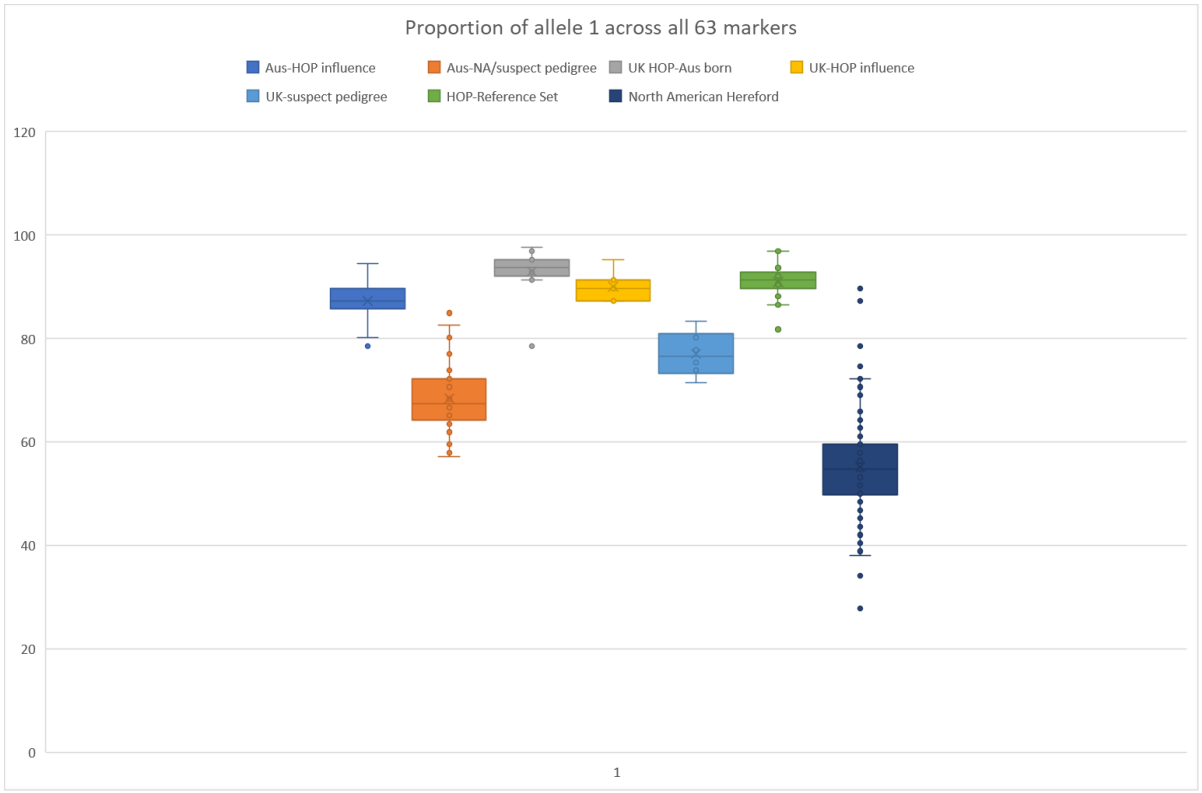
CHR	SNP	BP	Freq_HOP	FREQ_NA	P	Genes nearby
13	ARS-BFGL-NGS-55024	30713329	0.1897	0.6522	1.64E-06	In FAM188A - associated with fertility traits and meat quality
13	ARS-BFGL-NGS-119227	35563916	0.05172	0.4565	1.14E-06	In MTPAP - associated with meat quality and fertility
13	ARS-BFGL-NGS-21065	56698060	0.1034	0.5652	4.03E-07	Near CDH4, associated with milk protein percentage. Also near to SYCP2 - involved in udder development
14	ARS-BFGL-BAC-20261	10271449	0.03571	0.4348	1.08E-06	In EFR3A - linked to milk production
14	UA-IFASA-5964	19629540	0.1724	0.6304	1.65E-06	Near HAS2 - linked to Johnes disease resistance. Also linked to fertility
14	ARS-BFGL-BAC-1180	22226073	0.1897	0.6522	1.64E-06	In SNTG1 - linked to metabolic body weight in cattle and longevity in Fleckvieh
16	BTA-39689-no-rs	62555582	0.1724	0.6739	2.02E-07	Near TOR1AIP1 and 2, in eQTL region for meat quality
17	ARS-BFGL-NGS-10560	9921157	0.05172	0.4565	1.14E-06	In NR3C2, linked to production traits in Nellore cattle
17	ARS-BFGL-NGS-7812	17714385	0.5345	0.06522	4.04E-07	Near SCOC (associated with body size - long bone length - in mice). Also near CLGN, associated with fertility
17	ARS-BFGL-NGS-30619	50194955	0.1552	0.6364	5.55E-07	Near TMEM132C, linked to heifer rebreeding in Nellores
18	ARS-BFGL-NGS-18307	892087	0	0.3478	1.05E-06	Near UQCRFS1, linked to age at 1st calving
18	ARS-BFGL-NGS-118676	934619	0	0.3409	1.47E-06	Near UQCRFS1, linked to age at 1st calving
18	ARS-BFGL-NGS-116891	36146356	0.1379	0.6087	5.39E-07	Near CDH3 - linked to fertility traits
18	ARS-BFGL-BAC-36979	53571368	0.03448	0.4783	9.56E-08	In OPA3 - linked to dilated cardiomyopathy in Holstein-descended breeds
19	Hapmap50966-BTA-45035	33895220	0.08621	0.5435	3.18E-07	In PIGL - linked to endometriosis
19	BTB-00755526	48239722	0	0.3478	1.05E-06	In TANC2, associated with feed efficiency in Charolais
21	ARS-BFGL-NGS-101699	14000867	0.06897	0.5435	8.34E-08	Near RGMA and CHD2, linked to mental traits - docility or epilepsy?
21	ARS-BFGL-NGS-119716	57765296	0.01724	0.3913	9.44E-07	In SLC24A4, linked to attachment width and final conformation score
21	Hapmap35820-SCAFFOLD55013_40998	69222948	0.08621	0.5435	3.18E-07	Near TRAF3 - linked to immune response
21	ARS-BFGL-NGS-91002	69853542	0.06897	0.6522	3.03E-10	In APOPT1 - no obvious link to physical traits
25	ARS-BFGL-NGS-20408	520071	0.01724	0.3913	9.44E-07	In RAB40C, possible link to domestication
25	ARS-BFGL-NGS-3069	10968180	0.1552	0.6087	1.60E-06	In SNX29, associated with growth traits in cattle and pigs, plus human breast milk fatty acid content

CHR = chromosome, SNP = marker ID, BP = base position, Freq_Hereford (OP) = frequency of allele 1 in Hereford (OP), Freq_Hereford (NAD) = frequency of allele 1 in Hereford (NAD), P = raw p-value

As can be seen, there are large shifts in allele frequency between the Hereford (OP) and the Hereford (NAD), often representing a move towards markers being fixed in the opposite direction in the Hereford (NAD). Significant shifts in the frequency of alleles are possible with intensive selective breeding given either highly heritable phenotypes to select from or the availability of DNA testing. But affecting so many loci on different chromosomes from a source population where the frequency of each selected allele starts at ~2-10% of animals and moves to 40-60% is unlikely without either significant inbreeding effects or introgression of allelic material from another breed where the alternate alleles are more common.

The Lents' herd being shifted due to line-breeding is confirmed due to this analysis – on average, 80% of the 63 markers significant in the Hereford (OP) vs Hereford (NAD) are also observed in the Anxiety 4th animals. However, where these are observed, the Anxiety 4th tend towards being completely fixed (no variation). This is consistent with QC data provided by GeneSeek/NeoGen which showed that heterozygosity (markers where the animal carries both alleles) is only ~15% on average in the Anxiety 4th samples compared to other animals (typically ~30-35%). The highly significant markers identified in the Hereford (OP) vs Hereford (NAD) comparison typically display a pattern where the frequency of allele 1 is ~2-10% whilst in Hereford (NAD) it is 40-60% or *vice-versa*. This gives a high probability that the low frequency alleles in HOP were either missing in Anxiety 4th and other Hereford (OP) exported to the US, or were further removed due to the line-breeding.

Can we use these markers alone as a genetic test for "Hereford (OP)/non-Hereford (OP)"? Assessing the percentages of allele 1 across all 63 markers in the Australian animals:



On average, it is clear that these markers replicate the ability of the cluster plots and the admixture testing to distinguish Hereford (OP) and Hereford (NAD) genetics, although in the case of the Australian-NA/suspect pedigree animals there is some overlap with the Australian-Hereford (OP) influence group. This may warrant a closer inspection of the pedigree of these animals, but it makes a case for employing more than one metric for assessing Hereford (OP) and Hereford (NAD) genetics. Given the ever-decreasing cost of SNP assays (the 50K assay currently employed by HHBI is now the same price per sample as the 7K test used in our carcass verification work),

there is no reason not to use as many markers as possible to avoid claims that insufficient marker coverage could lead to mis-assignments (such claims have been levelled at the earlier work of Ogden as he only employed 96 SNPs, but as shown here this may have been ample for the question he was asked to address).

2.5. Sample Hereford (OP) verification reports

To demonstrate what a verification test for Hereford (OP) might look like, we present a three-way test based on the metrics described above – cluster position, admixture assignment and proportion of "Hereford (OP)" alleles across the 63 most significant markers.

Firstly, taking an animal (Baytal Charles) from the Baytal herd known to be of Hereford (OP) type:

Position on C1vC4 plot:

0.0937865, -0.0947975 (below -0.02 on C4)

Admixture assignments:

HOP-UK: 0.99988	US-Mark Donald: 0.00001
HOP-Australia: 0.00001	Canadian Influence: 0.00001
US-Line 1: 0.00001	All other breeds: 0.00008

Overall assignment to Hereford (OP) groupings = 0.99989 or >99.9%

Overall assignment to Hereford (NAD) groupings = 0.00003 or <0.1%

Proportion of Hereford (OP) allele type across 63 markers: 87%

Assessment: Hereford (OP)

Next, we take BT Trackman 585T, an animal known to come from corrupt genetics:

Position on C1vC4 plot:

0.0338713, 0.0339672 (above -0.02 on C4)

Admixture assignments:

HOP-UK: 0.00001	US-Mark Donald: 0.204588
HOP-Australia: 0.109055	Canadian Influence: 0.317638
US-Line 1: 0.088411	All other breeds: 0.280298

Overall assignment to Hereford (OP) groupings = 0.109065 or 10.9%

Overall assignment to Hereford (NAD) groupings = 0.610637 or 61.1%

Proportion of Hereford (OP) allele type across 63 markers: 56%

Assessment: Not Hereford (OP), strong evidence of hybrid genetics

Next let's look at an animal representative of the Hereford (NAD) based on most average C1C4 position for that grouping – individual Selkirk Lad H3:

Position on C1vC4 plot:

0.0854602, 0.0107256 (above -0.02 on C4)

Admixture assignments:

HOP-UK: 0.093023	US-Mark Donald: 0.417145
HOP-Australia: 0.172133	Canadian Influence: 0.251933
US-Line 1: 0.065685	All other breeds: 0.00008

Overall assignment to Hereford (OP) groupings = 0.265156 or 26.5%

Overall assignment to Hereford (NAD) groupings = 0.734763 or 73.5%

Proportion of Hereford (OP) allele type across 63 markers: 56%

Assessment: Not Hereford (OP), Hereford (NAD)

So what would an Australian animal of known HOP pedigree look like? Let's examine a representative animal – 393 Big Top Dowky 1:

Position on C1vC4 plot:

0.0613582, -0.0569802 (below -0.02 on C4)

Admixture assignments:

HOP-UK: 0.149193	US-Mark Donald: 0.028303
HOP-Australia: 0.757983	Canadian Influence: 0.00001
US-Line 1: 0.042098	All other breeds: 0.022413

Overall assignment to Hereford (OP) groupings = 0.907176 or 90.7%

Overall assignment to Hereford (NAD) groupings = 0.070411 or 7%

Proportion of Hereford (OP) allele type across 63 markers: 87%

Assessment: Hereford (OP)

Finally, what would an Australian animal of suspect or known NAD pedigree look like? Again, taking the most average animal from this grouping – Valma Odyssey:

Position on C1vC4 plot:

0.0474468, -0.00553797 (above -0.02 on C4)

Admixture assignments:

HOP-UK: 0.118886	US-Mark Donald: 0.149945
HOP-Australia: 0.261551	Canadian Influence: 0.220162
US-Line 1: 0.067816	All other breeds: 0.18164

Overall assignment to Hereford (OP) groupings = 0.380437 or 38%

Overall assignment to Hereford (NAD) groupings = 0.437923 or 43.8%

Proportion of Hereford (OP) allele type across 63 markers: 69%

Assessment: Not Hereford (OP), likely hybrid of NAD and additional breeds

2.6. What went into the Hereford (NAD)? Recommendations for future work.

The case for considering the Hereford (OP) and Hereford (NAD) two separate breeds, as both Ogden and Taylor proposed, is clear. The Hereford (NAD) is related to the Hereford (OP) by blood, but there has been considerable divergence highly unlikely to result from simple genetic drift or changes in breeding practice. Ogden and others have suspected the introgression of other breeds into the North American population, and the accounts given of the origin of the Polled Hereford certainly suggest that other breeds were used to create that particular grouping. Can we then determine what breeds were used in the formation of the Hereford (NAD)? Unfortunately, with the level of data used in this project, the answer is no – this test works because large sections of the genome become “fixed” in closed groups of cattle as they are selected for within breeds, becoming “signatures” of the breed. Where crossbreeding has occurred, these signatures are broken down. Whilst we can spot crossbred and quarterbred individuals (as we were asked to do in our earlier carcass project), as the animals are further crossbred or backcrossed to Hereford, this becomes more challenging.

However, with sufficient information, it may be possible – services are available for human genetics to tell you your percentage ancestry. This was done by identifying

sections of the genome that are unique to particular populations. For this to be replicated in the Hereford, we would need to use whole-genome sequencing of key individuals identified via this work as exemplars of Hereford (OP), Hereford (NAD) and admixed genetics. These can be used to identify sections of the genome that are unique to the Hereford (OP) and those unique to Hereford (NAD) types. The latter can then be compared to data from other cattle breeds (i.e. Simmental, Angus, Shorthorn) to see if they match. From these, a selection of genetic markers can be developed to add to future Hereford testing. With continuing support, it is our intention to conduct this further research into the Hereford genome.

3. Literature Cited:

Alexander D., Novembre J., Lange K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, **19**:1655–1664.

Decker, J. McKay, S., Rolf, M., Kim, J., Alcalá, A., Sonstegard T., *et al.* (2014) Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle. *PLoS Genetics*, https://doi.org/10.1371/journal.pgen.1004254.

Heath-Agnew, E. (1983) *A History of Hereford Cattle And Their Breeders.* Gerald Duckworth & Co. Ltd. (London).

MacDonald, J. Sinclair, J. (1909) *History of Hereford Cattle*. Vinton & Company Ltd. (London).

McMahon, R., Avila, V., Debbonaire, A., Patten, E., Muth, P. & Hegarty, M., (2015) Verification of breed, individual identity and farm of origin of beef carcasses using genetic assays. Report from Aberystwyth University to Waitrose and Dovecote Park Farms Ltd, available on request.

Murray-Dickson, G. (2014) Development of breed specific assays for beef and pork authentication. http://randd.defra.gov.uk/Document.aspx?Document=13075_FinalReportFA0125.pdf

Ogden, R. (2012) Extension of the method to verify meat from traditional cattle and pig breeds using SNP DNA markers - FA0112. http://sciencesearch.defra.gov.uk/Document.aspx?Document=11250_BreedsextensionFA0112-Finalreport.pdf

Ornduff, D. (1957) *The Hereford in America: A Compilation of Historic Facts About the Breed's Background and Bloodlines.* Self-published (Kansas City, Missouri).

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M., Bender, D., *et al.* (2007) PLINK: A Tool Set for Whole-Genome and Population-Based Linkage Analyses. *American Journal of Human Genetics*, 81.